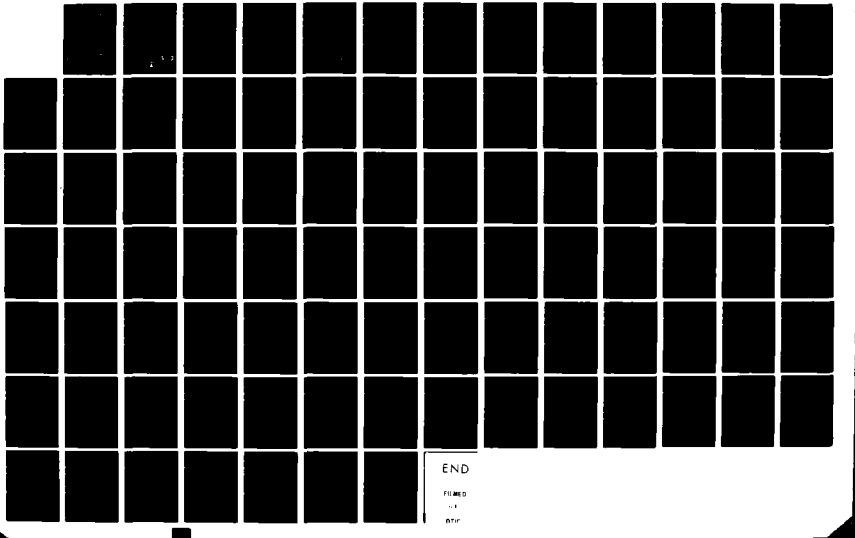
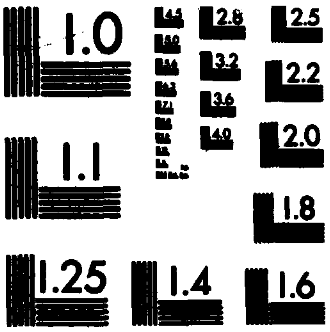
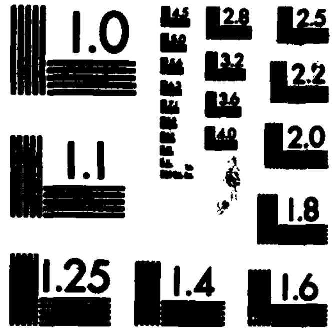
AD-A120 646 LARGE SAMPLE THEORY FOR SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL(U) STANFORD UNIV CA DEPT OF STATISTICS T SELLKE AUG 82 TR-20 N00014-77-C-0306 1/1

UNCLASSIFIED F/G 12/1 NL

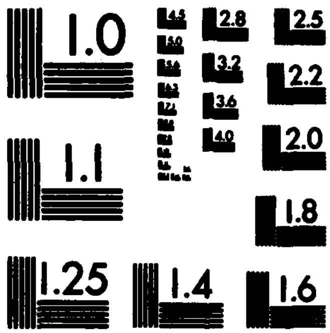




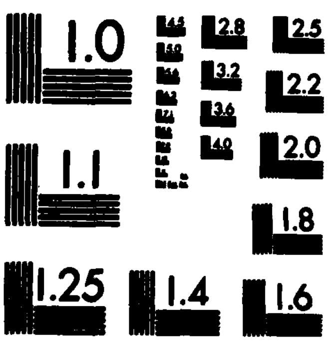
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



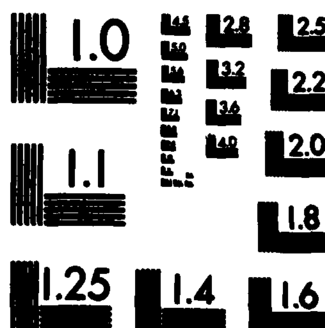
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

LARGE SAMPLE THEORY FOR SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL

BY

THOMAS SELLKE

TECHNICAL REPORT NO. 20
AUGUST 1982

PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

LARGE SAMPLE THEORY FOR SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL

by

Thomas Sellke
Stanford University

TECHNICAL REPORT NO. 20

August 1982

Prepared Under the Auspices
of
Office of Naval Research Contract
N00014-77-C-0306 (NR-042-373)

DEPARTMENT OF STATISTICS
Stanford University
Stanford, California

TABLE OF CONTENTS

SUMMARY

An appropriate large sample theory for sequential analysis of the Cox proportional hazards model is developed. For clinical trials with simultaneous entry of patients, the efficient score process of the partial likelihood is easily seen to be a martingale. It follows that, in a time scale based on the observed Fisher information, the score process and the properly normalized maximum partial likelihood estimator behave asymptotically like Brownian motion. When entry is staggered, the efficient score process is no longer a martingale in general. However, if patients in a staggered-entry clinical trial are assumed to be independent and identically distributed, independently of entry time, then the score process is well approximated by a martingale. The asymptotic results involving weak convergence to Brownian motion hold as before.

Key words: Proportional hazards model, sequential analysis.

CHAPTER I
INTRODUCTION AND SUMMARY

The proportional hazards model of survival analysis and its analysis by the method of partial likelihood originate in the work of Cox (1972, 1975), who argued that under general conditions maximum partial likelihood estimators have asymptotically normal distributions very similar to the asymptotic distributions of ordinary maximum likelihood estimators. The heuristic arguments given by Cox, though intuitively compelling, were nonrigorous and somewhat vague, and since then a number of authors have attempted to rigorously justify Cox's approach. See, for example, Bailey (1979), Gill (1980), Tsiatis (1981a), and Andersen and Gill (1981). For the most part, the work of these authors can be thought of as referring to medical trials in which all patients enter simultaneously or to medical trials with staggered entry in which statistical analysis is only carried out at a single predetermined time. However, the patients in a medical trial typically do not enter simultaneously, and uncertainty about the rate at which information will be accumulated in a clinical trial often makes the idea of choosing a termination time for the trial in advance rather dubious. In addition, there are the usual ethical and decision-theoretic arguments for analyzing the data from a medical trial sequentially so that the trial may be terminated quickly if large treatment effects appear to be present (see Armitage (1975)).

The goal of this dissertation is to justify sequential methods for statistical analysis of the proportional hazards model. Chapter II will deal with medical trials with simultaneous entry of patients. Although this context is often unrealistic, the results of Chapter II show that sequential analysis involving a random rescaling of time based on the observed Fisher information is a natural approach to the proportional hazards model. Chapter III obtains results like those of Chapter II for the case of staggered entry when the covariate and censoring characteristics of different patients are i.i.d., independently of entry time. The Appendix reviews several basic facts concerning martingales and proves a Skorokhod embedding theorem for martingales which is used in Chapters II and III.

Chapter I will proceed as follows. In Section 1.1, the proportional hazards model, the partial likelihood, and Cox's large sample theory arguments will be reviewed. The Bayesian view of the problem will also be considered. Section 1.2 will discuss the approach of Chapter II and compare it to the approach used by Andersen and Gill (1981) in a similar setting. Section 1.3 will summarize Chapter III and compare the results to those of Tsiatis (1981b) and Slud (1982), who also deal with sequential procedures for medical trials with staggered entry.

1.1. A Review of the Proportional Hazards Model

Suppose a medical trial involving n patients is conducted. Let $\lambda_i(\cdot)$ be the hazard rate for the i-th patient, so that

$$(1.1)\quad \lambda_i(t) = \lim_{h \to 0+} \frac{1}{h} \cdot \frac{P\{\text{patient } i \text{ dies in } [t, t+h)\}}{P\{\text{patient } i \text{ does not die before time } t\}} .$$

The Cox proportional hazards model assumes that the hazard rate for the i-th patient has the form

$$\lambda_i(t) = \lambda_0(t) \exp(\beta' z_i) \tag{1.2}$$

for $t \geq 0$. Here, $\lambda_0(\cdot)$ is a baseline hazard rate, and t is the elapsed time after entry into the medical trial. The components of the p-vector z_i are observable covariate values for the i-th patient, and β is a p-vector of unknown parameters. In general, z_i is permitted to vary with time, but in this chapter we will assume for simplicity that the z_i's are constant in time. The model allows (right) censoring of patients, with the obvious restriction that the censoring not "anticipate" deaths. The goal is to do statistical inference on β so as to determine how the covariate vector z_i influences the hazard rate. The unknown function $\lambda_0(\cdot)$ is regarded here as an infinite-dimensional nuisance parameter.

Suppose first that all n patients enter the trial simultaneously and that we have observed the trial over the time interval $[0,t]$. Let $x_i(t)$ be the amount of time patient i was under observation prior to possible censoring or death. Suppose $m = m(t)$ deaths were observed at times $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$, where patient (j) is the patient observed to die at time $t_{(j)}$. The results of such a medical trial up to time t may be summarized as in Figure 1.1. Define the risk set R_j of the j-th death by

$$R_j = \{i: x_i(t) \geq t_{(j)}\} . \tag{1.3}$$

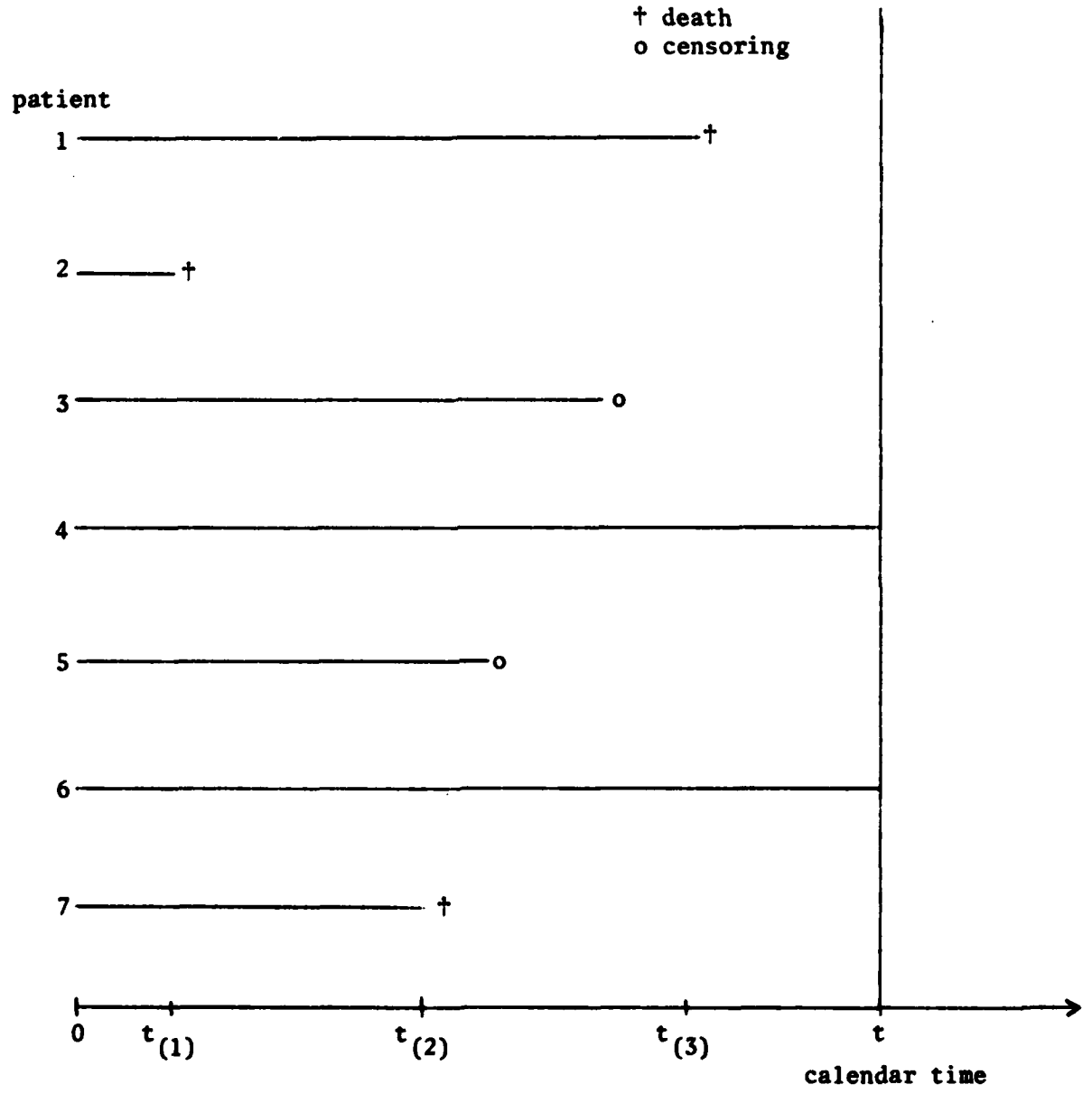


Figure 1.1. Survival times for a clinical trial with simultaneous entry of patients. Patients 4 and 6 are still in the trial at time t.

Thus, R_j consists of those patients who were in the trial and therefore at risk just prior to the j-th observed death.

Let us now review the definition and the rationale for Cox's partial likelihood. Let $\mathcal{H}t_{(j)-}$ represent the observations in the time interval $[0, t_{(j)})$ plus the occurrence of a death at time $t_{(j)}$. Let $\mathcal{H}t_{(j)}$ represent the observations in $[0, t_{(j)}]$. Thus, $\mathcal{H}t_{(j)}$ includes the identity of the patient (j), while $\mathcal{H}t_{(j)-}$ does not. For notational convenience, set $t_{(0)} = 0$ and $t_{(m+1)} = t$. Then Cox decomposes the likelihood function of the observations in $[0,t]$ as follows, where the probabilities[1] depend on $\lambda_0(\cdot)$, β, and the distributions of the censoring times.

$$(1.4) \qquad P\{\mathcal{H}t_{(m+1)}\} = \prod_{j=1}^{m+1} P\{\mathcal{H}t_{(j)} | \mathcal{H}t_{(j-1)}\}$$

$$= \prod_{k=1}^{m+1} P\{\mathcal{H}t_{(k)-} | \mathcal{H}t_{(k-1)}\} \cdot \prod_{j=1}^{m} P\{\mathcal{H}t_{(j)} | \mathcal{H}t_{(j)-}\} .$$

Note that

$$P\{\mathcal{H}t_{(k)-} | \mathcal{H}t_{(k-1)}\}$$

is the conditional probability of the event "no deaths in $(t_{(k-1)}, t_{(k)})$, observed censoring occurs in $(t_{(k-1)}, t_{(k)})$, and a death occurs at time $t_{(k)}$", given the observations in $[0, t_{(k-1)}]$. If one writes out this conditional probability in terms of β, $\lambda_0(\cdot)$, and the distributions of the censoring times, one gets a messy

[1] The term "probability" is being used rather loosely here, since these "probabilities" may include values of density functions

expression from which it does not seem possible to extract much information about β as long as the function $\lambda_0(\cdot)$ remains unknown. The factors in the second product are much more useful.

$$(1.5)\quad P\{Ht_{(j)} | Ht_{(j)^-}\} = P\{\text{patient (j) dies at time } t_{(j)} | R_j \text{ and a death occurs at time } t_{(j)}\}$$

$$= \frac{\exp\{\beta' z_{(j)}\}}{\sum_{i \in R_j} \exp\{\beta' z_i\}} .$$

The factors $\lambda_0(t_{(j)})$ have been cancelled out of (1.5), so that β is the only unknown remaining in (1.5). This second product in the likelihood (1.4)

$$(1.6)\qquad PL(t,\beta) = \prod_{j=1}^{m} \frac{e^{\beta' z_{(j)}}}{\sum_{i \in R_j} e^{\beta' z_i}}$$

is Cox's partial likelihood. Cox (1975) argues that one should ignore the other factor of the likelihood and use this partial likelihood in the same way that one uses an ordinary likelihood. Use of the partial likelihood is supported by Efron (1977), who shows that all but a small part of the information about β is typically contained in the partial likelihood even when the parametric form of $\lambda_0(\cdot)$ is known. Thus, the loss of information about β caused by use of the partial likelihood should be quite negligible when $\lambda_0(\cdot)$ is completely unknown. Moreover, use of the partial likelihood in the case where

no prior knowledge or only vague prior knowledge about $\lambda_0(\cdot)$ is available seems unavoidable if the problem is to be tractable.

Under the proportional hazards model, one can view such a medical trial as a random series of experiments. An experiment consists of having "nature" choose one person to die from the current risk set, where the probability that a given patient is chosen is proportional to this patient's value of $e^{\beta' z_i}$. The experiments themselves are random in that the risk sets in question and the death times are random. The partial likelihood is the combined likelihood for the outcomes of these experiments and ignores the randomness involved in the determination of which experiments are performed and when they are performed. The general situation of which this is a special case is a sequence of random experiments, where the conditional probability distribution of the outcome of an experiment, given previous observations, only depends on a parameter β, but where the choice of the next experiment to be performed is random, perhaps with dependence on the results of previous experiments, on β, and on nuisance parameters. The use of a partial likelihood for such sequences of random experiments is the subject of Cox (1975). Although the discussion that follows will refer to the proportional hazards model, the results for medical trials with simultaneous entry of patients will apply directly to the general situation.

For a Bayesian who is willing to use the partial likelihood, the problem of evaluating data from a medical trial is easy under this model. The Bayesian just multiplies his prior density for β by the partial likelihood, and the normalized product becomes his posterior density for β. Use of the partial likelihood avoids the

need to specify a prior distribution for $\lambda_0(\cdot)$. The Bayesian may still be faced with the decision of when to stop a medical trial, but this involves matters, such as loss functions and the cost of sampling, which will not be explicitly considered here.

Cox (1975) suggested that the usual large-sample theory for ordinary likelihoods be applied to the partial likelihood. The logarithm of the partial likelihood is given by

$$\ell(t,\beta) = \log PL(t,\beta) \tag{1.7}$$

$$= \sum_{j=1}^{m(t)} \{\beta' z_{(j)} - \log \sum_{i \in R_j} e^{\beta' z_i}\} .$$

For one-dimensional β and z_i, the efficient score is given by

$$\dot{\ell}(t,\beta) = \frac{\partial}{\partial \beta} \log PL(t,\beta) \tag{1.8}$$

$$= \sum_{j=1}^{m(t)} \{z_{(j)} - A_{(j)}(\beta)\}$$

where

$$A_{(j)}(\beta) = \frac{\sum_{i \in R_j} z_i e^{\beta' z_i}}{\sum_{i \in R_j} e^{\beta' z_i}} . \tag{1.9}$$

The observed Fisher information of the partial likelihood is given by

$$-\ddot{\ell}(t,\beta) = - \frac{\partial^2}{\partial \beta^2} \log PL(t,\beta) \tag{1.10}$$

$$= \sum_{j=1}^{m(t)} V_{(j)}(\beta)$$

where

$$V_{(j)}(\beta) = \frac{\sum_{i \in R_j} \{z_i - A_{(j)}(\beta)\}^2 e^{\beta' z_i}}{\sum_{i \in R_j} e^{\beta' z_i}} . \tag{1.11}$$

Note that $A_{(j)}(\beta)$ is the weighted average value of z_i in R_j, where each patient is weighted proportionally to $e^{\beta z_i}$. Thus, the increments of $\dot{\ell}(t,\beta)$ are equal to

$$z_{(j)} - E_\beta(z_{(j)}|R_j) . \tag{1.12}$$

It is also easy to see that

$$V_{(j)}(\beta) = \mathrm{var}_\beta(z_{(j)}|R_j) . \tag{1.13}$$

Cox (1975) claims that for large m

$$\dot{\ell}(t,\beta_0) \{-\ddot{\ell}(t,\beta_0)\}^{-\frac{1}{2}}$$

has approximately a $N(0,1)$ distribution when β_0 is the true value of β. Furthermore, "under weak conditions on the third derivative of the log likelihood", he claims that

$$(\hat{\beta}-\beta_0) \{-\ddot{\ell}(t,\beta_0)\}^{\frac{1}{2}}$$

has approximately a $N(0,1)$ distribution, where $\hat{\beta}$ is the maximum partial likelihood estimate of β_0. These results immediately generalize to the case of p-dimensional β and z_i. The efficient score

$$\dot{\ell}(t,\beta) = \nabla_\beta \log PL(t,\beta)$$

is now a $p\times 1$ vector, and the observed Fisher information

$$-\ddot{\ell}(t,\beta) = (\nabla_\beta)^2 \log PL(t,\beta)$$

is now a $p\times p$ matrix. The formulas (1.8) and (1.9) for $\dot{\ell}(t,\beta)$ remain valid, and (1.10) remains valid if the square $\{z_i - A_{(j)}(\beta)\}^2$ in (1.11) is replaced by the $p\times p$ matrix

$$\{z_i - A_{(j)}(\beta)\}\{z_i - A_{(j)}(\beta)\}^T .$$

Formula (1.12) still holds, and (1.13) becomes

$$V_{(j)}(\beta) = \mathrm{cov}_\beta(z_{(j)}|R_j) . \tag{1.14}$$

The asymptotic distributional results are the same with $N(0,1)$ replaced by the p-dimensional standard normal distribution $N(\underset{\sim}{0}, I_{p\times p})$.

In Cox's heuristic argument for the asymptotic distributions, he treats the number of deaths m as constant. The increments (1.12) are shown to be uncorrelated and with mean 0. He also assumes "some degree of independence" between the increments (1.12) of $\dot{\ell}(t,\beta)$ and that the increments $V_{(j)}(\beta)$ of $-\ddot{\ell}(t,\beta)$ are "not too disparate" in size. Under these somewhat vague conditions, the central limit theorem presumably applies.

1.2. Application of Martingale Theory to the Simultaneous Entry Case

It is not hard to see martingales lurking in the background in Cox's (1975) argument, and it seems to have become generally accepted

that martingale theory is the natural mathematical setting in which to investigate the large-sample theory of the partial likelihood. (cf. Aalen (1977, 1978, 1980), Gill (1980), and Andersen and Gill (1981).) Each coordinate of the efficient score process $\dot{\ell}(t,\beta)$ evaluated at the true value of β is easily seen to be a martingale in t, where the σ-algebra F_t is generated by events in $[0,t]$. The information process $-\ddot{\ell}(t,\beta)$ is the sum of the conditional (i.e., given $Ft_{(j)-}$) covariances $V_{(j)}(\beta)$ of the increments of $\dot{\ell}(t,\beta)$. Thus, the asymptotic normality of the efficient score $\dot{\ell}(t,\beta)$ is largely a question of whether a martingale central limit theorem applies. The basic requirements for applicability of a martingale central limit theorem to $\dot{\ell}(T,\beta)$, where T is a stopping time, are that $-\ddot{\ell}(T,\beta)$ be approximately constant (or, if random, approximately independent of the martingale process) and that the jumps of $\dot{\ell}(t,\beta)$, $t \leq T$, be small compared to $\{-\ddot{\ell}(T,\beta)\}^{\frac{1}{2}}$. The second requirement is a Lindeberg condition.

At least two approaches for ensuring the applicability of a martingale central limit theorem to the efficient score process $\dot{\ell}(t,\beta)$ are possible. The first, used by Andersen and Gill (1981), requires that the observed information matrix $-\ddot{\ell}(t,\beta)$ grow in an essentially nonrandom way. To be specific, let $I(\cdot)$, $0 \leq t \leq 1$, be a fixed, continuous $p \times p$ matrix valued function which is non-decreasing in the sense that $I(t_2) - I(t_1)$ is nonnegative definite for $t_2 > t_1$, and for which $I(1)$ is positive definite. Suppose that for each $n=1, 2, 3, \ldots$, we have a medical trial involving n patients, all of whom enter the trial at time $t=0$. The assumptions in Andersen and Gill (1981) imply that, as $n \to \infty$,

$$\sup_{t\in[0,1]} \| I(t) - n^{-1}\{-\ddot{\ell}_n(t,\beta)\} \| \xrightarrow{p} 0 ,$$

where $-\ddot{\ell}_n(t,\beta)$ is the observed Fisher information process for the n-th trial. This result, together with a Lindeberg condition, implies that $n^{-\frac{1}{2}} \dot{\ell}_n(t,\beta)$ converges weakly to a p-dimensional, independent increments Gaussian process with mean 0 and covariance matrix $I(t)$ at time t. If, in addition,

$$n^{-1}\{-\ddot{\ell}_n(1,\beta^*)\} \xrightarrow{p} I(1)$$

whenever β^* is a consistent estimator of β, then

$$n^{\frac{1}{2}}\{\hat{\beta}(1) - \beta\} \to N(\underset{\sim}{0}, I^{-1}(1)) ,$$

where $\hat{\beta}(1)$ is the maximum partial likelihood estimator of β at t=1. Andersen and Gill show that their assumptions are generally satisfied when the patients are independent and identically distributed with respect to covariates and censoring.

Suppose that z_i and β are one-dimensional. The approach used in Chapter II of this dissertation for guaranteeing the applicability of a martingale central limit theorem in the simultaneous entry case is to use the accumulated information $-\ddot{\ell}(t,\beta)$ as a clock time. Thus, by definition, information accumulates at a constant rate in this clock time. All that is needed in addition is a Lindeberg condition implying that the jumps of the $\dot{\ell}(t,\beta)$ and $-\ddot{\ell}(t,\beta)$ processes are not too big. Chapter II will assume that the covariates z_i are bounded in absolute value by a fixed constant B, so that the

necessary Lindeberg condition is trivially satisfied. It follows that, in the information time induced by $-\ddot{\ell}(t,\beta)$, the efficient score process $\dot{\ell}(t,\beta)$ and the properly normalized maximum partial likelihood estimator process

$$\{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}} \{\hat{\beta}(t) - \beta\}$$

are well approximated by a Brownian motion, and the approximation is shown to be uniformly good for medical trials satisfying $|z_i| \leq B$. If one is willing to work in this information time, the problem of sequentially estimating β or testing a hypothesis $H_0: \beta = \beta_0$ becomes asymptotically equivalent to sequentially estimating or testing the drift of a Brownian motion. The major disadvantage of this approach is that it does not generalize to the case of multidimensional z_i and β, since a $p \times p$ information matrix cannot be used as a clock time. However, the results of Chapter II illustrate that it is natural to operate in a clock time measuring information when the rate at which information will accumulate is unknown and perhaps random.

1.3. Medical Trials with Staggered Entry of i.i.d. Patients

Suppose that we have a medical trial like the one described in Section 1.1, except that the i-th patient now enters the trial at time y_i, where $0 \leq y_1 \leq y_2 \leq \ldots \leq y_n$. The observations of such a trial up until time t may be summarized graphically as in Figure 1.2. Let $x_i(t)$ again be the amount of time patient i was on test before time t prior to possible censoring or death. Let

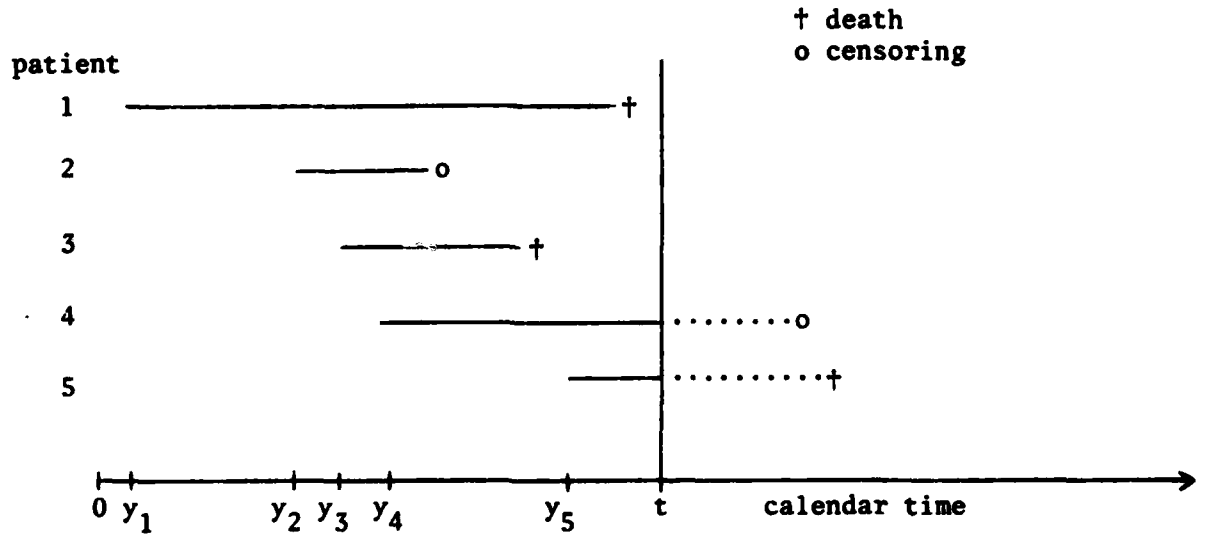


Figure 1.2. Survival times for a clinical trial with staggered entry. The dotted lines indicate what will be observed if the trial is continued after time t.

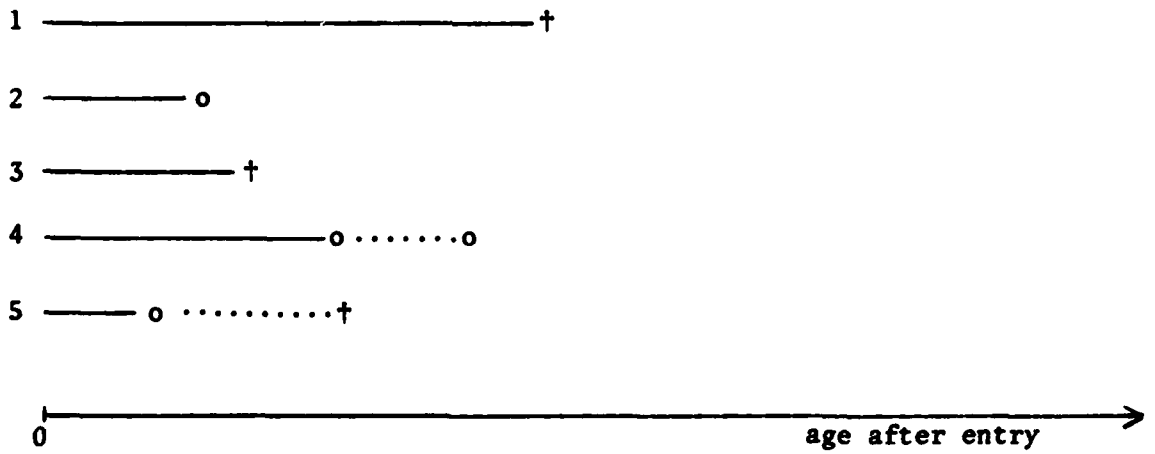


Figure 1.3. Survival times for an equivalent trial with simultaneous entry. Patients 4 and 5 are treated as censored when PL(t) is computed. Note that the risk set for the death of patient 3 changes if the original trial of Figure 1.2 is continued beyond time t.

$\Delta_i(t)$ equal 1 if patient i was observed to die before time t, and let $\Delta_i(t)$ equal 0 otherwise. Note that $x_i(t)$ cannot be greater than $(t - y_i)^+$, so that at time t patient i is in effect censored at age $(t - y_i)^+$. If we assume that the entry times are constant or have a distribution not depending on β, then the only data from patient i relevant to inference about β is the triple $\{z_i, x_i(t), \Delta_i(t)\}$. Thus, a simultaneous entry medical trial with the same values of $\{z_i, x_i(t), \Delta_i(t)\}$ as the trial described above would give the same information about β. The simultaneous entry trial of Figure 1.3 is equivalent to the staggered entry trial of Figure 1.2. The partial likelihood for staggered entry data is defined to equal the partial likelihood for the equivalent simultaneous entry data. Define

$$R(t,s) = \{i: x_i(t) \geq s\} , \tag{1.14}$$

so that $R(t,s)$ is the set of patients who by time t have been on test for at least s time units. If $\Delta_i(t) = 1$, then $R\{t, x_i(t)\}$ is the risk set at time t for the death of patient i. Note that the risk set for a death can now vary with t. The efficient score of the partial likelihood is given by

$$\dot{\ell}(t,\beta) = \sum_{i=1}^{n} [z_i - \tilde{\mu}_\beta\{t, x_i(t)\}] \Delta_i(t) , \tag{1.15}$$

where

$$\tilde{\mu}_\beta(t,s) = \frac{\sum_{j \in R(t,s)} z_j e^{\beta' z_j}}{\sum_{j \in R(t,s)} e^{\beta' z_j}} . \tag{1.16}$$

The observed Fisher information of the partial likelihood is

$$-\ddot{\ell}(t,\beta) = \sum_{i=1}^{n} \tilde{\sigma}^2\{t, x_i(t)\}\, \Delta_i(t) \tag{1.17}$$

where

$$\tilde{\sigma}^2_\beta(t,s) = \frac{\sum_{j\in R(t,s)} [z_j - \tilde{\mu}_\beta(t,s)][z_j - \tilde{\mu}_\beta(t,s)]^T e^{\beta' z_j}}{\sum_{j\in R(t,s)} e^{\beta' z_j}}. \tag{1.18}$$

Note that $\tilde{\mu}_\beta(t,s)$ is the weighted average of covariates for patients in $R(t,s)$, and $\tilde{\sigma}^2_\beta(t,s)$ is the covariance matrix for this weighted distribution of covariate vectors.

If we are only going to analyze data from a staggered entry trial at a single fixed time t, then the above discussion shows that the problem is equivalent to analyzing data from a simultaneous entry trial at a single time. The results of Andersen and Gill (1981) described in the last section are very useful in this case. A Bayesian would also see little difference between simultaneous entry and staggered entry, since the procedure for updating a prior density for β would be the same in either case. However, the behavior of $\dot{\ell}(t,\beta)$ and of $\hat{\beta}(t)$ as processes in t is much more difficult to analyze under staggered entry, since $\dot{\ell}(t,\beta)$ is not generally a martingale. The efficient score $\dot{\ell}(t,\beta)$ is still a sum over observed deaths of the difference between the covariate of the dying patient and the weighted average of covariates in the risk set, but the risk set no longer consists of those patients who were on

test at the time of death, and the risk sets even change with t. Thus, the interpretation of the trial as a sequence of random experiments no longer makes sense.

Although Jones and Whitehead (1979) proposed a sequential test for staggered entry trials and did a computer simulation of their procedure, the first theoretical result on the joint distribution of $\dot{\ell}(t,\beta)$ is due to Tsiatis (1981b). Tsiatis assumes that n patients have i.i.d. entry times distributed on a finite interval. The patients also have i.i.d. one-dimensional covariates which are constant in time and independent of the entry times. Censoring is not allowed. If $n\to\infty$, with the entry time distribution, the covariate distribution, and the hazard function held fixed, Tsiatis shows that, under the null hypothesis H_0: $\beta=0$, the joint distribution of $n^{-\frac{1}{2}}\,\dot{\ell}(t,0)$ at fixed times $t_1, t_2, \ldots, t_k$ converges in law to a multivariate normal distribution with mean 0 and independent increments. The asymptotic variance of $n^{-\frac{1}{2}}\,\dot{\ell}(t_i, 0)$ is shown to be proportional to $P\{\Delta_1(t_i) = 1\}$. The proof is based on a clever decomposition of $\dot{\ell}(t,0)$, which, in the notation of Chapter III, can be written as

$$\dot{\ell}(t,0) = Q(t) + r(t) \,. \tag{1.19}$$

The $Q(t)$ process is a sum of n i.i.d. martingales, one for each patient. The term $r(t)$ is shown to be $o(n^{\frac{1}{2}})$ in probability for fixed t. The result follows easily from the multivariate central limit theorem. There are several weakenesses in Tsiatis' result. On the one hand, the assumptions are *extremely* strong. On the other

hand, the conclusion is rather weak and not well suited to some applications. The theorem specifies that the times $t_1, \ldots, t_k$ must be fixed. Thus, they must be chosen in advance, even though one may have little knowledge of how fast information will accumulate from the trial. Tsiatis proposes that the times be chosen so that

$$P\{\Delta_1(t_i) = 1\} = i/k, \quad i=1, 2, \ldots, k , \tag{1.20}$$

but this would demand good prior knowledge about both the entry time distribution and the hazard rate.

Chapter III of this dissertation proves results for staggered entry trials very similar to those given in Chapter II for simultaneous entry trials. Again, in the information time induced by $-\ddot{\ell}(t,\beta)$, the score process $\dot{\ell}(t,\beta)$ and the normalized maximum partial likelihood estimator process

$$\{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}} \{\hat{\beta}(t) - \beta\}$$

are well approximated by a Brownian motion. The conditions are much weaker than those of Tsiatis. As in Chapter II, the covariates are assumed to be one-dimensional and bounded in absolute value by a fixed constant B. The z_i's are allowed to vary with time, and right censoring is permitted. The central distributional assumption is that the patients are i.i.d. with respect to covariates and censoring, independently of entry times. The entry times themselves are treated as ancillary.

The proof of the theorem in Chapter III is long and technical, but it can be split into three parts:

(a) The efficient score process $\dot{\ell}(t,\beta)$ is shown to be close to a martingale $Q(t)$. The observed Fisher information process $-\ddot{\ell}(t,\beta)$ is shown to be close to $\langle Q \rangle(t)$, the predictable quadratic variation process of $Q(t)$. The proofs use a generalized version of (1.19).

(b) Consistency of $\hat{\beta}(t)$ as an estimator of β follows from part (a). A Taylor series argument shows that $\{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}} \{\hat{\beta}(t) - \beta\}$ is close to $\dot{\ell}(t,\beta)$.

(c) The martingale $Q(t)$ is well approximated by a Brownian motion in the clock time induced by $\langle Q \rangle(t)$. The proof uses a Skorokhod embedding theorem for martingales proved in the Appendix.

Putting (a) and (c) together implies that $\dot{\ell}(t,\beta)$ is well approximated by a Brownian motion in information time. By (b), the same holds for $\{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}} \{\hat{\beta}(t) - \beta\}$. Subject to an innocuous technical condition, the approximation by Brownian motion is uniformly good for trials satisfying $|z_i| \leq B$ and with β in a given compact interval. The final section of Chapter III discusses the generalization of these results to the multivariate case.

The recent manuscript of Slud (1982) also attempts to show that the efficient score process under staggered entry converges weakly to a Brownian motion in a suitable time scale. Slud introduces a martingale which is similar to our $Q(t)$ to approximate

the score process. He considers only $\beta=0$ and uses a time renormalization which would be inappropriate for general β. Also, what corresponds to our Proposition 3.1 is essentially his assumption A.5. This assumption is never actually verified, although Slud states that it can be verified under various sets of conditions, all of which require strong hypotheses on the arrival process.

CHAPTER II

TRIALS WITH SIMULTANEOUS ENTRY OF PATIENTS

The purpose of this chapter is to show that asymptotic normality of the maximum partial likelihood estimator holds in great generality when the following three requirements are met.

(2.1a) The covariate processes $z_i(\cdot)$ are one-dimensional.

(2.1b) The hazard rate for the i-th patient has the form $\lambda_0(t) \exp(\beta z_i(t))$, *where* t *is now the calendar time after the beginning of the trial.*

(2.1c) The estimation is done sequentially in a clock time (called information time) measuring observed Fisher information.

When conditions (2.1a) and (2.1b) are satisfied, the efficient score process is generally a one-dimensional martingale. The information time of (2.1c) is equal to the sum of the conditional variances of the jumps of this efficient score martingale. Weak convergence of the efficient score process in this information time to standard Brownian motion follows from a martingale central limit theorem. A Taylor series argument shows that the maximum partial likelihood estimator process in information time also looks like a Brownian motion when the estimator process is properly normalized.

If the hazard rate for patients in a medical trial is a function of covariates and of time after entry into the trial, then (2.1b) demands that all patients enter simultaneously. Alternatively, (2.1b) is satisfied if the hazard rate for an individual under study depends on covariates and on varying environmental influences which at any

time affect all individuals at risk equally. An example would be the following model for the occurrence of auto accidents among the people driving in a given area. (Since the version of the proportional hazards model used here implies that simultaneous events do not occur, one could eliminate simultaneous accidents by saying that a multiple vehicle accident only counts against the driver who is most at fault.) The intensity of the Poisson-like accident process for a person who is driving at time t could be assumed to have the form $\lambda_0(t) \exp(\beta z(t))$, where $z(t)$ gives the level of alcohol in the driver's blood, and $\lambda_0(t)$ depends on environmental conditions to which all people driving in the area at time t are equally subject, such as weather, time of day, and traffic conditions. At any time t, the risk set would be the set of all people actually driving in the area. A person not driving at time t is considered to be censored. Since it is possible for a person to cause more than one accident and since the censoring process is more complicated than the usual simple right censoring, it would be necessary to use the counting process formulation of the Cox model found in Andersen and Gill (1981) and described below.

2.1. Notation and Formulation of the Model

As was indicated above, the method of this chapter requires that the efficient score process be a one-dimensional martingale. A rather general formulation of the Cox model which satisfies this requirement when the covariates are one-dimensional is given in Andersen and Gill (1981).

Following Andersen and Gill (1981), suppose we have a medical trial involving n patients, to each of whom there corresponds a counting process $N_i(\cdot)$ which counts observed events in the life of the i-th individual. These events could be deaths, so that at most one event is observed for each individual, or they could be recurrent events, such as epileptic seizures or outbreaks of a rash. The sample functions $N_i(\cdot)$ look like those of a Poisson process in that they are increasing, right-continuous step functions which increase by jumps of size $+1$ and satisfy $N_i(0) = 0$. It is required that $N_i(\cdot)$ and $N_j(\cdot)$ do not jump simultaneously for $i \neq j$.

Let $\{\Omega, \mathcal{F}, P\}$ be the probability space on which the stochastic processes $N_i(\cdot)$ are defined for $t \geq 0$. Let $\mathcal{F}_t$ be the σ-algebra generated by everything that happens in $[0,t]$, so that $\{\mathcal{F}_t, t \in [0,\infty]\}$ is a right-continuous, nondecreasing family of σ-algebras. It is assumed that $E\, N_i(t) < \infty$ for $t < \infty$ and that the counting process $N_i(\cdot)$ has a random intensity function

$$\lambda_i(t) = C_i(t)\, \lambda_0(t)\, e^{\beta z_i(t)}, \qquad t \geq 0 . \tag{2.2}$$

Here β is a scalar parameter, $z_i(\cdot)$ is the one-dimensional covariate process for patient i, and $\lambda_0(\cdot)$ is a fixed baseline hazard function. The censoring process $C_i(\cdot)$ equals 1 when patient i is under observation and is 0 otherwise. The processes $z_i(\cdot)$ and $C_i(\cdot)$ are assumed to be $\mathcal{F}_t$-predictable. This condition is satisfied if these processes are adapted and left-continuous with right-hand limits. It will also be assumed that the $z_i(\cdot)$ processes are bounded in absolute value by a fixed constant B. The situation described here will be referred to as a B-experiment, and the asymptotic results of

this chapter will be found to hold uniformly in B-experiments. The assumption that (2.1) is the intensity process for $N_i(\cdot)$ means that

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(u)du \tag{2.2}$$

is a local $\mathcal{F}_t$-martingale with predictable quadratic variation

$$\langle M_i(t),\, M_i(t)\rangle = \int_0^t \lambda_i(u)du \tag{2.3}$$

and that $M_i(\cdot)$ and $M_j(\cdot)$ are orthogonal martingales for $i \neq j$.

In this setting, the logarithm of the partial likelihood can be written

$$\log PL(t,\beta) = \sum_i \int_0^t [\beta z_i(s) - \log\{\sum_j C_j(s)\, e^{\beta z_j(s)}\}]\, dN_i(s)\ . \tag{2.4}$$

If we take the partial derivative of (2.4) with respect to β, we see that the efficient score process for the partial likelihood is given by

$$\dot{\ell}(t,\beta) = \sum_i \int_0^t \{z_i(s) - \tilde{\mu}(s)\}\, dN_i(s)\ , \tag{2.5}$$

where

$$\tilde{\mu}(s) = \frac{\sum_j C_j(s)\, z_j(s)\, e^{\beta z_j(s)}}{\sum_j C_j(s)\, e^{\beta z_j(s)}}\ . \tag{2.6}$$

In (2.6) and elsewhere, we interpret $0/0$ as 0. It follows from algebra that

$$\dot{\ell}(t,\beta) = \sum_i \int_0^t \{z_i(s) - \tilde{\mu}(s)\}\, dM_i(s)\ . \tag{2.7}$$

The integrands $z_i(s) - \tilde{\mu}(s)$ are bounded, $\mathcal{F}_s$-predictable functions. It follows from the theory of stochastic integrals (see Gill (1980), p. 10) that $\dot{\ell}(t,\beta)$ is a local martingale with predictable quadratic variation

$$(2.8) \qquad \left\langle \dot{\ell}(t,\beta),\ \dot{\ell}(t,\beta) \right\rangle = \sum_i \int_0^t \{z_i(s) - \tilde{\mu}(s)\}^2 \lambda_i(s) ds \ .$$

Andersen and Gill (1981) proceed to impose conditions which imply that, for some fixed function $I(\cdot)$ on $[0,1]$,

$$n^{-1} \sup_{t\in[0,1]} \left| \left\langle \dot{\ell}(t,\beta),\ \dot{\ell}(t,\beta) \right\rangle - I(t) \right| \xrightarrow{p} 0$$

as the number of patients $n\to\infty$. They are able to conclude from the martingale central limit theorem of Rebolledo (1980) that $n^{-\frac{1}{2}} \dot{\ell}(\cdot,\beta)$ converges weakly as $n\to\infty$ to a Gaussian independent-increments process on $[0,1]$ with variance function $I(\cdot)$. Andersen and Gill use multi-dimensional $z_i(\cdot)$ processes without the assumption that the covariates are bounded, but the basic idea is as described here.

It will be useful in this chapter if we introduce a new family of σ-algebras. First define the ordered event times $t_{(1)} < t_{(2)} < t_{(3)} < \ldots$, and let $t_{(m)} = \infty$ if fewer than m events occur. Also take $t_{(0)} = 0$. Now define

$$(2.9) \qquad \mathcal{F}_k^+ = \sigma\{t_{(k+1)};\ \mathcal{F}_t,\ t < t_{(k+1)}\} \ , \qquad k=0,\ 1,\ 2,\ \ldots \ ,$$

and $\mathcal{F}_\infty^+ = \mathcal{F}_\infty$. Thus, $\mathcal{F}_k^+$ represents everything that happens until just prior to the $(k+1)$st observed event, together with knowledge of when, if ever, the $(k+1)$st event occurs.

Define

$$Y_k = \dot{\ell}(t_{(k)}, \beta) , \qquad k=0, 1, 2, \ldots \tag{2.10}$$

$$X_k = Y_k - Y_{k-1} , \qquad k=1, 2, \ldots \tag{2.11}$$

and

$$v_k = \mathrm{var}(X_k | F_{k-1}^+) , \qquad k=1, 2, \ldots . \tag{2.12}$$

Then, at least under sufficient regularity and probably in general, $\{Y_k, F_k^+\}_{k=0}^{\infty}$ is a martingale for which the conditional variances v_k of the martingale differences X_k satisfy

$$\sum_{i=1}^{k} v_i = -\ddot{\ell}(t_{(k)}, \beta) . \tag{2.13}$$

Note that

$$-\ddot{\ell}(t,\beta) = \sum_i \int_0^t \frac{\sum_j C_j(s)\{z_j(s) - \tilde{\mu}(s)\}^2 e^{\beta z_j(s)}}{\sum_j C_j(s) e^{\beta z_j(s)}} dN_i(s) . \tag{2.14}$$

The heuristic argument is as in Section 1.1: conditional on F_{k-1}^+, the k-th event consists of nature randomly choosing a patient out of the risk set at time $t_{(k)}$ with probabilities proportional to the weights $e^{\beta z_j(t_{(k)})}$. Thus, X_k is the difference between the covariate of the patient chosen and the weighted average $\tilde{\mu}(s)$ of covariates, so that $E(X_k | F_{k-1}^+) = 0$. Furthermore, v_k is the variance of the covariates in the weighted distribution, so that (2.13) holds.

2.2. Approximation of the Score Process by a Brownian Motion

For $u \in [0,\infty)$, define

$$k(u) = \sup\{k: \sum_{i=1}^{k} v_i \leq u\} \tag{2.15}$$

$$= \sup\{k: -\ddot{\ell}(t_{(k)}, \beta) \leq u\} .$$

Now define the information time version of the efficient score process for $u \in [0,\infty)$, by

$$S(u) = Y_{k(u)} \tag{2.16}$$

$$= \dot{\ell}(t_{(k(u))}, \beta) .$$

Also define $\mathcal{G}_u = \mathcal{F}^+_{k(u)}$, $u \in [0,\infty)$, and

$$T = -\ddot{\ell}(\infty,\beta) \tag{2.17}$$

$$= \sum_{i=1}^{\infty} v_i .$$

Since v_i is $\mathcal{F}^+_{i-1}$ measurable, $k(u)$ is a stopping time with respect to the family of σ-algebras $\{\mathcal{F}^+_k, k=0, 1, \ldots, \infty\}$. Hence, it is easy to show via the martingale convergence theorem that $S(\cdot)$ is a $\mathcal{G}_u$-martingale for $u \in [0,\infty)$. Also, T is a $\mathcal{G}_u$ stopping time since

$$\{T \leq u\} = \bigcap_n \{\sum_1^n v_i \leq u\} = \bigcap_n \{k(u) \geq n\} \in \mathcal{G}_u .$$

The $\mathcal{G}_u$-martingale $S(\cdot)$ is seen to remain constant for $u \geq T$. By (2.13), the predictable quadratic variation process of $S(\cdot)$ is $-\ddot{\ell}(t_{(k(\cdot))}, \beta)$, so that by (2.15) the predictable quadratic variation of $S(\cdot)$ at time u is between $u - B^2$ and u.

Suppose we have a sequence of B-experiments indexed by m, $m=1, 2, \ldots$. Suppose further that

$$P\{T^{(m)} > m\} \to 1 \quad \text{as} \quad m \to \infty . \tag{2.18}$$

Then it follows immediately from the martingale central limit theorem of Rebolledo (1980) (see Theorem A.1 of the Appendix) that, as $m \to \infty$,

$$m^{-\frac{1}{2}} S^{(m)}((\cdot)m) \xrightarrow{\mathcal{L}} W(\cdot) \tag{2.19}$$

on $[0,1]$, where $W(\cdot)$ is a standard Brownian motion.

However, it seems more enlightening to approximate $S(\cdot)$ directly by a Brownian motion. To do this, it may be necessary to enlarge the probability space $\{\Omega, F, P\}$. Let $A_1, A_2, \ldots$ be independent random variables distributed uniformly on $[0,1]$ and defined on another probability space $\{X, A, \mu\}$. Let $\{\Omega^*, F^*, P^*\} = \{\Omega \times X, F \times A, P \times \mu\}$ be the product probability space. Define $F_k^{**} = F_k^+ \times A_{2k}$, where $A_k = \sigma\{A_1, \ldots, A_k\}$. The starred random variables Y_k^*, v_k^*, $S_{(u)}^*$, etc. are just Y_k, v_k, $S(u)$, etc. considered as random variables on the product space. The following proposition is an immediate consequence of Theorem A.2 in the Appendix.

Proposition 2.1 (Skorokhod representation for $\{Y_k\}$.)

There exists a standard Brownian motion $W(\cdot)$ and a sequence of random variables $0 = \tau_0 \le \tau_1 \le \tau_2 \le \ldots$ on $\{\Omega^*, F^*, P^*\}$ such that (2.20) holds.

$$X_k^* = W(\tau_k) - W(\tau_{k-1}) \tag{2.20a}$$

$$E(\tau_k - \tau_{k-1} | F_{k-1}^{**}) = v_k^* \tag{2.20b}$$

(2.20c) $$\mathrm{var}(\tau_k - \tau_{k-1} | F_{k-1}^{++}) \le 2B^2 v_k^*$$

(2.20d) τ_k is F_k^{++}-measurable, and the pre-τ_k σ-algebra of $W(\cdot)$ is contained in F_k^{++}.

Theorem 2.2 (Approximation of the efficient score process by a Brownian motion)

Let $W(\cdot)$ be the Brownian motion of Proposition 2.1, and let $\epsilon > 0$. Then, as $K \to \infty$,

$$P\{|S^*(u) - W(u)| < K + u^{\frac{1}{4}+\epsilon}, \quad \forall u \in [0,T]\} \to 1 ,$$

uniformly in B-experiments.

Proof. By (2.20a) and the definition of $S^*(\cdot)$,

(2.21) $$S^*(u) = W(\tau_k) \quad \text{for} \quad \sum_1^k v_i^* \le u < \sum_1^{k+1} v_i^* .$$

The idea here is to show that $\tau_k - \sum_1^k v_i^*$ is sufficiently small so that $W(u) - W(\tau_k)$ is small for u between $\sum_1^k v_i^*$ and $\sum_1^{k+1} v_i^*$.

By (2.20b,d), $\tau_k - \sum_1^k v_i^*$ is an F_k^{++}-martingale. By (2.20c) and Kolmogorov's inequality (see Doob (1953), Theorem 3.2, p. 314), for each $m=1, 2, \ldots$ we have

(2.22) $$P\{\sup_{k \le k(2^m)} |\tau_k - \sum_1^k v_i^*| \ge 2^{m(1+\epsilon)/2}\} \le \frac{2B^2}{2^{m\epsilon}} .$$

Using the formula

$$P\{\sup_{u \le a} |W(u)| > b\} \le 4\{1 - \Phi(\frac{b}{\sqrt{a}})\}$$

yields

$$(2.23)\qquad P\{\sup_{\substack{0\le u\le 2^m \\ 0\le h\le 2^{m(1+\varepsilon)/2}}} \tfrac{1}{2}|W(u+h) - W(u)| > 2^{\frac{m}{4}+\frac{m\varepsilon}{2}}\}$$

$$\le P\{\sup_{\substack{0\le j\le 2^{m/2} \\ 0\le h\le 2^{m(1+\varepsilon)/2}}} |W(u_j+h) - W(u_j)| > 2^{\frac{m}{4}+\frac{m\varepsilon}{2}}\}$$

$$\le 2^{m/2} \cdot 4 \cdot \{1 - \Phi(2^{m\varepsilon/4})\} ,$$

where

$$u_j = j2^{m/2} \text{ for } j=0, 1, \ldots .$$

Since

$$\sum_{m=1}^{\infty} \frac{2B^2}{2^{m\varepsilon}} < \infty \quad \text{and} \quad \sum_{m=1}^{\infty} 2^{m/2}\{1 - \Phi(2^{m\varepsilon/4})\} < \infty ,$$

it follows from (2.21), (2.22), and (2.23) that

$$(2.24)\qquad P\{|S^*(u) - W(u)| < u^{\frac{1}{4}+\varepsilon},\ u \in [L,T]\} \to 1$$

as $L\to\infty$, uniformly in B-experiments. Another application of Kolmogorov's inequality to $S^*(u)$ for $u \le L$ finishes the proof.

In terms of the observable processes $\dot{\ell}(t,\beta)$ and $-\ddot{\ell}(t,\beta)$, Theorem 2.2 says that, as $K\to\infty$,

$$(2.25)\quad P[|\dot{\ell}(t,\beta) - W\{-\ddot{\ell}(t,\beta)\}| < K + \{-\ddot{\ell}(t,\beta)\}^{\frac{1}{4}+\varepsilon},\ \forall\ t \ge 0] \to 1 ,$$

uniformly in B-experiments. Thus $\dot{\ell}(t,\beta)$ looks like a standard Brownian motion in the time scale determined by $-\ddot{\ell}(t,\beta)$. This result can be used to construct sequential tests of known size of the

hypothesis H_0: $\beta = \beta_0$. However, the theorem in the next section is needed if we wish to calculate the power of a test or to sequentially estimate β.

2.3. Approximation of the Maximum Partial Likelihood Estimator Process by a Brownian Motion

Let $\hat{\beta}(t)$ be the maximum partial likelihood estimator of β at time t. Let $W(\cdot)$ be the Brownian motion of Proposition 2.1.

Theorem 2.3. Let $\varepsilon > 0$. Then, as $L \to \infty$,

$$P\{\left|[-\ddot{\ell}\{t, \hat{\beta}(t)\}] \cdot \{\hat{\beta}(t) - \beta\} - W[-\ddot{\ell}\{t, \hat{\beta}(t)\}]\right| < [-\ddot{\ell}\{t, \hat{\beta}(t)\}]^{\frac{1}{4}+\varepsilon}$$

$$\text{for } -\ddot{\ell}\{t, \hat{\beta}(t)\} > L\} \to 1$$

uniformly in B-experiments.

The proof of Theorem 2.3 has been omitted since it is very similar to Section 3.4 of the next chapter.

Theorem 2.3 says that if $\hat{\beta}(t) - \beta$ is normalized by multiplication by $[-\ddot{\ell}\{t, \hat{\beta}(t)\}]$, the resulting process looks like a standard Brownian motion in the time scale determined by the estimated observed Fisher information $-\ddot{\ell}\{t, \hat{\beta}(t)\}$. The observable process $[-\ddot{\ell}\{t, \hat{\beta}(t)] \cdot \hat{\beta}(t)$ looks like a Brownian motion with drift β in the $-\ddot{\ell}\{t, \hat{\beta}(t)\}$ time scale. Thus, the problem of sequentially estimating β or testing H_0: $\beta = \beta_0$ has been shown to be asymptotically equivalent to the problem of sequentially estimating or testing the drift of a Brownian motion. There are two remaining difficulties. The first is that $-\ddot{\ell}\{\infty, \hat{\beta}(\infty)\}$ may be smaller than we would like, so that the experiment stops giving additional information before we choose to stop sampling.

In terms of the Brownian motion problem, this would mean that we might not be able to watch the Brownian motion for as long as we would like to. However, if $-\ddot{\ell}\{\infty, \hat{\beta}(\infty)\} < \infty$, then (2.20d) of Proposition 2.1 suggests that the behavior of

$$\dot{\ell}(t,\beta)$$

and of

$$[-\ddot{\ell}\{t, \hat{\beta}(t)\}]\ \{\hat{\beta}(t) - \beta\}$$

in information time is approximately that of a Brownian motion until a stopping time. Thus, the information time at which the experiment ends does not somehow anticipate what the score process or the maximum partial likelihood estimator process would have done if they had been allowed to continue.

The other difficulty results from the fact that $-\ddot{\ell}\{t, \hat{\beta}(t)\}$ may grow very slowly when $|\beta|$ is very large. To be specific, suppose we are to compare two treatments for a disease, so that z_i is either 0 or 1, depending on the treatment group. We wish to conduct a sequential test of $\beta=0$ by observing the process $[-\ddot{\ell}\{t, \hat{\beta}(t)\}] \cdot \hat{\beta}(t)$ in the $-\ddot{\ell}\{t, \hat{\beta}(t)\}$ time scale. If we intend to use the Brownian motion approximation to compute the power function and expected information time until stopping for our sequential test, then Theorem 2.3 says that we should continue the trial at least until $-\ddot{\ell}\{t, \hat{\beta}(t)\} > L$ for a value of L chosen to make the Brownian motion approximation good. However, if β is very large, we may see many deaths in treatment group 1 before we see any deaths in

treatment group 0. If all observed deaths are from treatment group 1, then $\hat{\beta}(t) = \infty$ and $-\ddot{\ell}\{t, \hat{\beta}(t)\} = 0$. Even if deaths have been observed in both groups, $-\ddot{\ell}\{t, \hat{\beta}(t)\}$ may be small long after one treatment has shown itself to be much better than the other. To avoid this problem, one could observe the $\beta=0$ score process $\dot{\ell}(t,0)$ in the $-\ddot{\ell}(t,0)$ time scale until $-\ddot{\ell}\{t, \hat{\beta}(t)\} \geq L$. If $|\dot{\ell}(t,0)|$ exceeds some number M before $-\ddot{\ell}\{t, \hat{\beta}(t)\} \geq L$, then we stop and reject H_0: $\beta=0$. Otherwise, we begin to observe $[-\ddot{\ell}\{t, \hat{\beta}(t)\}]\,\hat{\beta}(t)$ in the $-\ddot{\ell}\{t, \hat{\beta}(t)\}$ time scale when $-\ddot{\ell}\{t, \hat{\beta}(t)\}$ exceeds L. If M is reasonably large, then the probability of stopping before $-\ddot{\ell}\{t, \hat{\beta}(t)\} \geq L$ is small for moderate values of β, and the power function and expected information time until stopping can be found from the Brownian motion approximation. For large values of $|\beta|$, this procedure does not force us to continue the medical trial long after common sense tells us to stop.

CHAPTER III

TRIALS WITH STAGGERED ENTRY AND INDEPENDENT IDENTICALLY DISTRIBUTED PATIENTS

In Chapter II it was seen from

$$\dot{\ell}(t,\beta) = \sum_i \int_0^t \{z_i(s) - \tilde{\mu}(s)\}\, dM_i(s) \tag{2.7}$$

that $\dot{\ell}(t,\beta)$ was a local martingale. In the case of staggered entry of patients, we shall obtain a similar equation

$$\dot{\ell}(t,\beta) = \sum_i \int_{[0,t]} \{z_i(s-y_i) - \tilde{\mu}(t,\, s-y_i)\}\, dM_i(s)\ , \tag{3.1}$$

where y_i is the entry time of the i-th patient (cf. equation 3.22)). The M_i's are still local martingales, but now the integrands are functions of t for each s. Let $R(t,s)$ be the set of patients who, by calendar time t, have been under observation for s time units after entry. Then $\tilde{\mu}(t,s)$ is the weighted average of covariates of patients in $R(t,s)$. One can still use the martingale property of M_i to study $\dot{\ell}(t,\beta)$ for a single, fixed value of t, but the $\dot{\ell}(t,\beta)$ process itself is no longer a martingale in general.

This chapter extends the results of Chapter II to the case of staggered entry. The basic idea is as follows. If we could replace the random function $\tilde{\mu}(t,\, s-y_i)$ appearing in (3.1) with a nonrandom function $\mu(s-y_i)$ not depending on t, then the resulting integral would be a martingale in t. In a time scale determined by its predictable quadratic variation, this martingale could be approximated by a Brownian motion as was done in Chapter II, and one would hope that this predictable quadratic variation process is well approximated by $-\ddot{\ell}(t,\beta)$. In order to guarantee that $\tilde{\mu}(t,s)$ is sufficiently well

approximated by a deterministic function $\mu(s)$, we will assume that the (one-dimensional) covariate processes and the censoring times of patients are jointly i.i.d., independently of entry times. Then the patients in $R(t,s)$ have conditionally i.i.d. covariates, where this conditional distribution does not depend on t. If the cardinality of $R(t,s)$ is large, then $\tilde{\mu}(t,s)$ is approximately nonrandom in the sense that its distribution is very concentrated around $\mu(s)$. In order to ensure that the risk sets $R(t,s)$ are large "most of the time", it will be necessary to abandon the very general counting process and censoring process formulation of Chapter II and to return to the more standard setting of "deaths" and right censoring.

3.1. Notation and Formulation of the Model

Suppose we are given a possibly infinite sequence of entry times $0 \leq y_1 \leq y_2 \leq \ldots$ such that any interval $[0,t]$ contains only finitely many y_i's. To the patient i entering at time y_i is associated a random triple $\{z_i(\cdot), c_i, h_i\}$. The one-dimensional covariate process $z_i(\cdot)$ defined on $[0,\infty)$ is assumed left-continuous and bounded in absolute value by a fixed constant B. The possibly infinite random variable c_i is the time after entry of censoring, and the random variable h_i is the amount of "accumulated hazard" which patient i can tolerate before dying. Also given is a fixed baseline cumulative hazard function $\Lambda(s)$, $s \geq 0$, which satisfies $\Lambda(0) = 0$, is nondecreasing, and is continuous on $[0, t^{\infty})$, where $t^{\infty} = \inf\{t: \Lambda(t) = \infty\}$. Fix $\beta \in \mathbb{R}$ and define

(3.2) $$x_i = \inf\{t: \int_0^t e^{\beta z_i(s)} \, d\Lambda(s) \geq h_i\} .$$

The random variable x_i is the survival time of patient i after entry into the trial. The i-th patient is on test during the time interval $[y_i, y_i + x_i \wedge c_i]$. If $s \leq x_i \wedge c_i$, then at time $y_i + s$ we observe the covariate value $z_i(s)$. At time $y_i + x_i \wedge c_i$, we observe the death of the i-th patient if $x_i \leq c_i$ and otherwise observe that he is censored. At any time t there is in effect a second censoring variable $(t - y_i)^+$ in the sense that the time on test of patient i prior to t is $x_i \wedge c_i \wedge (t-y_i)^+$. We shall refer to x_i, c_i, and $(t - y_i)^+$ as "age" variables - the age of the i-th patient at death, at censoring, and at time t, respectively.

Our stochastic assumptions are as follows. The pairs $\{z_i(\cdot), c_i\}$ are independent and identically distributed, independently of the arrival times. The "tolerance to hazard" random variables h_i are exponentially distributed with parameter 1, independently of each other and of everything else. It will also be assumed that there exist positive numbers δ and η such that

(3.3) $$\Lambda(\delta) > 0$$

and

(3.4) $$\mathrm{var}\{z_1(s) \mid x_1 \wedge c_1 \geq s\} \geq \eta^2 \quad \text{for} \quad 0 \leq s \leq \delta .$$

A medical trial as described above will be referred to as a (B,δ,η)-experiment. A medical trial which satisfies all of the above conditions except possibly (3.3) and (3.4) will be referred to as a

B-experiment. Some of the asymptotic results of this chapter will hold uniformly in B-experiments, and all will hold uniformly in (B,δ,η)-experiments.

All probabilities and expectations should be considered as conditional, given $y_1, y_2, \ldots$.

It is convenient to introduce the notation

$$N_i(t,s) = I_{\{y_i+x_i\le t,\ x_i\le c_i,\ x_i\le s\}}\ , \tag{3.5}$$

to indicate that the i-th patient arrived and died before time t, and that he was uncensored and of age $\le s$ at the time of death. We also define the set of patients at risk at time t and age s by

$$R(t,s) = \{i:\ y_i \le t-s,\ x_i \wedge c_i \ge s\}\ . \tag{3.6}$$

With this notation Cox's (1975) log partial likelihood for β can be expressed by

$$\ell(t,\beta) = \sum_i \int_{[0,t]} [\beta z_i(s) - \log\{\sum_{j\in R(t,s)} e^{\beta z_j(s)}\}]\ N_i(t,ds)\ . \tag{3.7}$$

Differentiating (3.7) with respect to β gives the score process

$$\dot{\ell}(t,\beta) = \sum_i \int_{[0,t]} \{z_i(s) - \tilde{\mu}(t,s)\}\ N_i(t,ds) \tag{3.8}$$

where

$$\tilde{\mu}(t,s) = \frac{\sum_{j\in R(t,s)} z_j(s)\, e^{\beta z_j(s)}}{\sum_{j\in R(t,s)} e^{\beta z_j(s)}} \tag{3.9}$$

Minus the second derivative of (3.7) is the observed Fisher information process

$$-\ddot{\ell}(t,\beta) = \sum_i \int_{[0,t]} \tilde{\sigma}^2(t,s)\, N_i(t,ds) \,, \tag{3.10}$$

where

$$\tilde{\sigma}^2(t,s) = \frac{\sum_{j\in R(t,s)} \{z_j(s) - \tilde{\mu}(t,s)\}^2 e^{\beta z_j(s)}}{\sum_{j\in R(t,s)} e^{\beta z_j(s)}} \tag{3.11}$$

$$= \frac{\sum_{j\in R(t,s)} z_j^2(s)\, e^{\beta z_j(s)}}{\sum_{j\in R(t,s)} e^{\beta z_j(s)}} - \tilde{\mu}^2(t,s) \,.$$

The maximum partial likelihood estimator of β is the solution $\hat{\beta} = \hat{\beta}(t)$ of

$$\dot{\ell}(t,\beta) = 0 \,.$$

Tests of the hypothesis $H_0\colon \beta = \beta_0$ can be based on $\hat{\beta}$ or directly on $\dot{\ell}(t,\beta_0)$. The usual Taylor series approximation

$$0 = \dot{\ell}(t,\hat{\beta}) = \dot{\ell}(t,\beta) + (\hat{\beta}-\beta)\, \ddot{\ell}(t,\ ) + \ldots \tag{3.12}$$

indicates that the asymptotic behavior of $\hat{\beta}$ is intimately associated with that of $\dot{\ell}(t,\beta)$, which we now consider.

Define

$$\Lambda_i(s) = \int_{[0,s]} e^{\beta z_i(u)}\, d\Lambda(u) \,. \tag{3.13}$$

For $s \geq 0$, let $\hat{\mathcal{F}}_s$ be the σ-algebra generated by y_i, c_i, $\{z_i(u), u \geq 0\}$, $I_{\{x_i \leq s\}}$, and $x_i I_{\{x_i \leq s\}}$, $i=1, 2, \ldots$. Then since

$$P\{x_i \in (s, s+\Delta] | \hat{\mathcal{F}}_s\} = \{1 - e^{\Lambda_i(s)-\Lambda_i(s+\Delta)}\} I_{\{x_i \geq s\}} \tag{3.14}$$

$$= [\{\Lambda_i(s+\Delta)-\lambda_i(s)\} + o\{\Lambda_i(s+\Delta)-\Lambda_i(s)\}] I_{\{x_i \geq s\}}$$

it follows that

$$I_{\{x_i \leq s\}} - \Lambda_i(x_i \wedge s) \tag{3.15}$$

is an $\hat{\mathcal{F}}_s$-martingale in $s \geq 0$. Fix $t \geq 0$. Since $c_i \wedge (t-y_i)^+$ is an $\hat{\mathcal{F}}_s$ stopping time,

$$I_{\{x_i \leq s \wedge c_i \wedge (t-y_i)^+\}} - \Lambda_i\{x_i \wedge s \wedge c_i \wedge (t-y_i)^+\} \tag{3.16}$$

is also an $\hat{\mathcal{F}}_s$-martingale. Let $\mathcal{F}_{t,s}$ be the sub-σ-algebra of $\hat{\mathcal{F}}_s$ containing events which have been observed by time t and which are of age $\leq s$, i.e. $\mathcal{F}_{t,s}$ is the σ-algebra generated by $I_{\{y_i \leq t\}}$, $y_i I_{\{y_i \leq t\}}$, $I_{\{x_i \leq s \wedge c_i \wedge (t-y_i)^+\}}$, $x_i I_{\{x_i \leq s \wedge c_i \wedge (t-y_i)^+\}}$, $I_{\{c_i \leq s \wedge x_i \wedge (t-y_i)^+\}}$, $c_i I_{\{c_i \leq s \wedge x_i \wedge (t-y_i)^+\}}$, and $\{z_i(u), u \in (0, s \wedge x_i \wedge c_i \wedge (t-y_i)^+]\}$, $i=1, 2, \ldots$. Since (3.16) is adapted to $\mathcal{F}_{t,s}$, it follows that (3.16) is also an $\mathcal{F}_{t,s}$-martingale in $s \geq 0$ for fixed t.

Let $A_i(t,ds) = I_{\{i \in R(t,s)\}} \Lambda_i(ds)$. Then

$$A_i(t,s) = \Lambda_i\{s \wedge x_i \wedge c_i \wedge (t-y_i)^+\} \tag{3.17}$$

and the martingale of (3.16) can be written as

(3.18) $$N_i(t,s) - A_i(t,s) \ .$$

Define

(3.19) $$\dot{\ell}(t,s,\beta) = \sum_i \int_{[0,s]} \{z_i(u) - \tilde{\mu}(t,u)\}\{N_i(t,du) - A_i(t,du)\} \ .$$

It follows from (3.8) and simple algebra that

$$\dot{\ell}(t,t,\beta) = \dot{\ell}(t,\beta) \ .$$

Moreover, the stochastic integral in (3.19) inherits the martingale property of (3.18) since the integrand is bounded and $\mathcal{F}_{t,s}$-predictable. Thus, for each fixed t,

(3.20) $$\{\dot{\ell}(t,s,\beta), \mathcal{F}_{t,s}\}$$

is a martingale in s (Gill (1980), p. 10 or Liptser and Shiryayev (1978), p. 268).

This martingale property in s of $\dot{\ell}(t,s,\beta)$ is the basis of the analysis of the asymptotic normality of $\dot{\ell}(t,\beta) = \dot{\ell}(t,t,\beta)$ at one fixed point in time found in Gill (1980) and in Andersen and Gill (1981). The idea is that if the observations up to time t are viewed in "age time" with patient i being censored at age time $c_i \wedge (t-y_i)^+$, then the age time process is equivalent to an experiment with simultaneous entry. However, this approach does not work if one is interested in the joint distribution of $\dot{\ell}(t,\beta)$ at different values of t since $\dot{\ell}(t,\beta)$ is not in general a martingale.

Let $N_i(t) = N_i(t,t)$, $A_i(t) = A_i(t,t)$, and $\mathcal{F}_t = \mathcal{F}_{t,t}$. Then $N_i(t)$ is an indicator for the event that patient i was observed

to die before time t, and $A_i(t)$ is the accumulated hazard to which patient i has been exposed by time t while under observation. An argument similar to the one showing that (3.18) is an $F_{t,s}$-martingale shows that

$$\{N_i(t) - A_i(t),\ F_t\} \tag{3.21}$$

is a martingale in t.

By a change of variable in (3.19),

$$\dot{\ell}(t,\beta) = \sum_i \int_{[0,t]} \{z_i(s-y_i) - \tilde{\mu}(t,\ s-y_i)\}\{N_i(ds) - A_i(ds)\} \ . \tag{3.22}$$

With the notation

$$M_i(s) = N_i(s) - A_i(s) \ ,$$

(3.22) is seen to be the same as (3.1). As was mentioned at the beginning of this chapter, the plan is to approximate $\tilde{\mu}(t,s)$ by a deterministic function $\mu(s)$. A natural candidate for $\mu(s)$ is given by

$$\mu(s) = \frac{E\{z_1(s)\ e^{\beta z_1(s)}\ ;\ x_1 \wedge c_1 \geq s\}}{E\{e^{\beta z_1(s)}\ ;\ x_1 \wedge c_1 \geq s\}} \ , \tag{3.23}$$

since an informal law of large numbers argument suggests that $\tilde{\mu}(t,s)$ should be close to $\mu(s)$ when $R(t,s)$ is large. Again, 0/0 is to be interpreted as 0 in (3.9), (3.11), (3.23), and elsewhere. Let

$$Q(t) = \sum_i \int_{[0,t]} \{z_i(s-y_i) - \mu(s-y_i)\}\{N_i(ds) - A_i(ds)\} \ . \tag{3.24}$$

Since the integrands in (3.24) are bounded and F_t-predictable,

$$\{Q(t),\ F_t\} \tag{3.25}$$

is a martingale in t. Note that $Q(t)$ can also be written as

$$Q(t) = \sum_i \int_{[0,t]} \{z_i(s) - \mu(s)\}\{N_i(t,ds) - A_i(t,ds)\} . \tag{3.26}$$

Define

$$N(t) = \sum_i N_i(t), \quad A(t) = \sum_i A_i(t), \quad \text{and} \quad D(t) = EN(t) . \tag{3.27}$$

Thus, $N(t)$ is the number of deaths observed by time t, $A(t)$ is the accumulated hazard acquired by all patients while on test before time t, and $D(t)$ is the expected number of deaths observed before time t. Also define

$$N(t,s) = \sum_i N_i(t,s) , \qquad A(t,s) = \sum_i A_i(t,s) \tag{3.28}$$

and

$$\begin{aligned} r(t) &= \dot{\ell}(t,\beta) - Q(t) \\ &= \sum_i \int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}\{N_i(t,ds) - A_i(t,ds)\} \\ &= \int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}\{N(t,ds) - A(t,ds)\} . \end{aligned} \tag{3.29}$$

The goal of this chapter is to show that $\dot{\ell}(t,\beta)$ and $[-\ddot{\ell}\{t, \hat{\beta}(t)\}] \cdot \{\hat{\beta}(t) - \beta\}$ can be well approximated by $W[-\ddot{\ell}\{t, \hat{\beta}(t)\}]$,

where $W(\cdot)$ is a standard Brownian motion. An argument similar to that of Chapter II shows that the martingale $Q(t)$ is well approximated by $W\{\langle Q\rangle(t)\}$, where $W(\cdot)$ is a Brownian motion and

$$\langle Q\rangle(t) = \sum_i \int_{[0,t]} \{z_i(s) - \mu(s)\}^2 A_i(t,ds) \tag{3.30}$$

is the predictable quadratic variation process of $Q(t)$. To apply this result to $\dot{\ell}(t,\beta)$, it is necessary to show that $r(t)$ given in (3.29) is small and that $-\ddot{\ell}(t,\beta)$ is close to $\langle Q\rangle(t)$. Some Taylor series arguments show that $\dot{\ell}(t,\beta)$ is close to $[-\ddot{\ell}\{t, \hat{\beta}(t)\}] \cdot \{\hat{\beta}(t) - \beta\}$.

Let $0 < \varepsilon < \frac{1}{15}$. The final theorem will follow from Propositions 3.1 through 3.6, all of which hold uniformly in (B,δ,η)- experiments, and for β in compact intervals. In addition, Propositions 3.1, 3.2, and 3.6 hold uniformly in B-experiments.

Proposition 3.1. As $K\to\infty$,

$$P\{|r(t)| \le K + D(t)^{\frac{1}{2}-\varepsilon},\ t \ge 0\} \to 1 ,$$

uniformly in B-experiments.

Proposition 3.2. As $K\to\infty$,

$$P\{|\ddot{\ell}(t,\beta) + \langle Q\rangle(t)| \le K + D(t)^{1-2\varepsilon},\ t \ge 0\} \to 1 ,$$

uniformly in B-experiments.

Proposition 3.3. There exists $\alpha > 0$ such that, as $K\to\infty$,

$$P\{\langle Q\rangle(t) + K > \alpha D(t),\ t \ge 0\} \to 1 .$$

Proposition 3.4. (Consistency of $\hat{\beta}(t)$.) As $L \to \infty$,

$$P[|\hat{\beta}(t) - \beta| < \{-\ddot{\ell}(t,\beta)\}^{\varepsilon - \frac{1}{2}} \text{ for } -\ddot{\ell}(t,\beta) > L] \to 1 .$$

Proposition 3.5. As $L \to \infty$,

$$P[|\ddot{\ell}\{t, \hat{\beta}(t)\}\{\hat{\beta}(t) - \beta\} + \dot{\ell}(t,\beta)| < \{-\ddot{\ell}(t,\beta)\}^{3\varepsilon} \text{ for } -\ddot{\ell}(t,\beta) > L] \to 1 .$$

Proposition 3.6. There exists a standard Brownian motion $W(\cdot)$ such that

$$P\{|Q(t) - W\{\langle Q \rangle(t)\}| < K + \langle Q \rangle(t)^{\frac{1}{4}+\varepsilon} \text{ for } t \geq 0\} \to 1 \text{ as } K \to \infty .$$

This holds uniformly in B-experiments.

Finally, we get to the main theorem.

Theorem 3.7. There exists a standard Brownian motion $W(\cdot)$ such that, as $K \to \infty$,

$$P[|\dot{\ell}(t,\beta) - W\{-\ddot{\ell}(t,\beta)\}| \leq K + \{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}-\varepsilon} \text{ for all } t \geq 0] \to 1 ,$$

and, as $L \to \infty$,

$$P[|[-\ddot{\ell}\{t, \hat{\beta}(t)\}] \cdot \{\hat{\beta}(t) - \beta\} - W[-\ddot{\ell}\{t, \hat{\beta}(t)\}]|$$

$$\leq [-\ddot{\ell}\{t, \hat{\beta}(t)\}]^{\frac{1}{2}-\varepsilon} \text{ for } -\ddot{\ell}\{t, \hat{\beta}(t)\} > L] \to 1 .$$

Furthermore, the convergence is uniform in (B,δ,η)-experiments, and for β in compact intervals.

3.2. Approximation of $\dot{\ell}(t,\beta)$ by the Martingale $Q(t)$

This section will show that

$$r(t) = \dot{\ell}(t,\beta) - Q(t)$$

is uniformly small in the sense of Proposition 3.1. The martingale property in s of $N_i(t,s) - A_i(t,s)$ for fixed t is used to show that

$$E\, r^2(t) = O\{3 + \log D(t)\} ,$$

uniformly in $t \in [0,\infty]$. The Chebyshev inequality is then applied to show that

$$|r(t_k)| \leq K + \tfrac{1}{2}D(t_k)^{\frac{1}{2}-\varepsilon}$$

holds with high probability for all t_k's in a certain sequence which increases to $+\infty$. Crude estimates which show that $r(t)$ does not vary too much between t_k's finish the proof of Proposition 3.1.

Lemma 3.8. For all $t \in [0,\infty]$,

$$E\, r^2(t) \leq 4B^2\, e^{4B|\beta|}\, \{3 + \log D(t)\} .$$

Proof. From fundamental properties of stochastic integrals,

$$E\, r^2(t) = E\Big[\sum_i \int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}^2\, N_i(t,ds)\Big] . \tag{3.31}$$

By considering the i-th term and conditioning on x_i, $R(t,x_i)$, and the event $\{x_i \leq c_i \wedge (t-y_i)\}$, we obtain

$$(3.32)\quad E\, r^2(t) = E\{\sum_i N_i(t,x_i)\, E[\{\tilde{\mu}(t,x_i) - \mu(x_i)\}^2 \mid x_i,\ R(t,x_i),\ x_i \le (t-y_i)^+ \wedge c_i]\}$$

$$\le e^{2B|\beta|}\, E\left(\sum_i \frac{N(t,x_i)}{|R(t,x_i)|^2}\, E\left[\left\{\mu_0(x_i) \sum_{j\in R(t,x_i)} z_j(x_i)\, e^{\beta z_j(x_i)} - \mu_1(x_i) \sum_{j\in R(t,x_i)} e^{\beta z_j(x_i)}\right\}^2 \,\middle|\, x_i,\ R(t,x_i),\ x_i \le (t-y_i)^+ \wedge c_i\right]\right),$$

where $\mu_\nu(s) = E\{z_1^\nu(s)\, e^{\beta z_1(s)} \mid x_1 \wedge c_1 \ge s\}$ for $\nu=0$ and 1, and $|A|$ denotes the cardinality of the set A. Let $R_i^*(t,s) = R(t,s) - \{i\}$, and observe that given x_i, $x_i \le c_i \wedge (t-y_i)^+$, and $R_i^*(t,x_i) = \{j_1, \ldots, j_m\}$, $z_{j_1}(x_i), \ldots, z_{j_m}(x_i)$ are independent and identically distributed with

$$E\left\{z_{j_\ell}^\nu(x_i)\, e^{\beta z_{j_\ell}(x_i)} \,\middle|\, x_i,\ R_i^*(t,x_i),\ x_i \le c_i \wedge (t-y_i)^+\right\} = \mu_\nu(x_i)$$

for $\nu=0$ or 1. Hence except for terms involving i, the conditional expectations on the right-hand side of (3.32) involve the square of a sum of i.i.d. random variables having mean 0 and variance less than $4B^2\, e^{2B|\beta|}$. Thus

$$(3.33)\quad E\, r^2(t) \le 4B^2\, e^{4B|\beta|}\, E\left\{\sum_i \frac{N_i(t,x_i)}{|R(t,x_i)|} + \sum_i \frac{N_i(t,x_i)}{|R(t,x_i)|^2}\right\}$$

$$\le 4B^2\, e^{4B|\beta|}\, E\{3 + \log N(t)\}$$

$$\le 4B^2\, e^{4B|\beta|}\, \{3 + \log D(t)\}\ .$$

The second to last inequality follows from

$$\sum_{i=1}^{N} \left(\frac{1}{i} + \frac{1}{i^2}\right) \leq \{3 + \log N\},$$

and the final inequality follows from the Jensen inequality. This finishes the proof of Lemma 3.8.

Let $0 < \varepsilon < \frac{1}{15}$, and define $0 = t_0 \leq t_1 \leq \ldots \leq \infty$ by

$$t_k = \inf\{t: D(t) = k^{1+3\varepsilon}\} . \tag{3.34}$$

Thus, $t_k = \infty$ if $D(\infty) \leq k^{1+3\varepsilon}$. By the Chebyshev inequality and Lemma 3.8,

$$P\{|r(t_k)| > \tfrac{1}{2} D^{\frac{1}{2}-\varepsilon}(t_k)\} \leq \text{const.} \frac{3 + \log D(t_k)}{D(t_k)^{1-2\varepsilon}} \tag{3.35}$$

$$\leq \text{const.} \frac{3 + (1+3\varepsilon) \log k}{k^{1+\varepsilon-6\varepsilon^2}}$$

for $t_k < \infty$. Since $1+\varepsilon-6\varepsilon^2 > 1$, the following lemma now follows easily.

Lemma 3.9. For $0 < \varepsilon < \frac{1}{15}$ and $\{t_k\}$ as above,

$$\lim_{n\to\infty} P\{|r(t_k)| \leq n^2 + \tfrac{1}{2} D^{\frac{1}{2}-\varepsilon}(t_k) \text{ for } k \geq n\} \to 1 .$$

The proof of Proposition 3.1 is completed by Lemmas 3.10 and 3.11, which show that $r(t)$ does not vary too much between 0 and t_n and between t_k and t_{k+1} for $k \geq n$, respectively.

Lemma 3.10. $P\{\max_{0\le t\le t_n} |r(t)| < n^2\} \to 1$ as $n\to\infty$.

Proof. Note that if $\mathfrak{B}(p)$ is a Bernoulli variable with parameter p, then for $p \le \frac{1}{2}$, $\mathfrak{B}(p)$ is stochastically smaller than a Poisson variable with parameter $2p$, say $\mathfrak{P}(2p)$. Hence for all $0 \le p \le 1$, $\mathfrak{B}(p)$ is stochastically smaller than $2p + \mathfrak{P}(2p)$. Since $N(t_n)$ is a sum of independent Bernoulli variables with $E\,N(t_n) \le n^{1+3\varepsilon}$, it follows that $N(t_n)$ is stochastically less than $2n^{1+3\varepsilon} + \mathfrak{P}(2n^{1+3\varepsilon})$. By the central limit theorem,

$$P\{N(t_n) > n^{1+4\varepsilon}\} \to 0 \quad \text{as} \quad n\to\infty \,. \tag{3.36}$$

On $\{A(t_n) \ge n^{1+5\varepsilon}\}$, $N(t_n)$ is stochastically larger than a Poisson random variable with mean $n^{1+5\varepsilon}$ and hence

$$P\{N(t_n) < n^{1+4\varepsilon},\ A(t_n) \ge n^{1+5\varepsilon}\} \le P\{\mathfrak{P}(n^{1+5\varepsilon}) \le n^{1+4\varepsilon}\} \to 0 \quad \text{as} \quad n\to\infty \,, \tag{3.37}$$

again by the central limit theorem.

By (3.36) and (3.37),

$$P\{A(t_n) > n^{1+5\varepsilon}\} \to 0 \quad \text{as} \quad n\to\infty \,. \tag{3.38}$$

By (3.29),

$$\max_{0\le t\le t_n} |r(t)| \le 2B\{N(t_n) + A(t_n)\} \,. \tag{3.39}$$

Together, (3.36), (3.38), and (3.39) imply Lemma 3.10.

Lemma 3.11.

$$P\{\max_{t_k \le t < t_{k+1}} |r(t) - r(t_k)| \le \frac{1}{2} D^{\frac{1}{2}-\epsilon}(t_k) \text{ for all } k \ge n\} \to 1$$

as $n \to \infty$.

Proof. Let $D_k = N(t_k) - N(t_{k-1})$ be the number of deaths observed in $(t_{k-1}, t_k]$. By (3.34)

$$E\, D_k \le k^{1+3\epsilon} - (k-1)^{1+3\epsilon} \le (1 + 3\epsilon)k^{3\epsilon} . \tag{3.40}$$

D_k is a sum of independent Bernoulli variables, so that by the argument in the proof of Lemma 3.10, D_k is stochastically smaller than $3k^{3\epsilon} + \mathcal{P}(3k^{3\epsilon})$. By easy large deviation estimates (see Feller (1971), p. 549, Theorem 1),

$$P\{D_k < k^{7\epsilon/2} \text{ for all } k \ge n\} \to 1 \text{ as } n \to \infty . \tag{3.41}$$

Let $H_k = A(t_k) - A(t_{k-1})$ be the hazard accumulated by patients while under observation in $(t_{k-1}, t_k]$. On $\{H_k \ge k^{4\epsilon}\}$, D_k is stochastically larger than a Poisson random variable with mean $k^{4\epsilon}$ and hence

$$P\{D_k < k^{7\epsilon/2}, H_k \ge k^{4\epsilon}\} \le P\{\mathcal{P}(k^{4\epsilon}) \le k^{7\epsilon/2}\} . \tag{3.42}$$

By (3.41), (3.42), and another easy large deviations estimate,

$$P\{H_k < k^{4\epsilon} \text{ for all } k \ge n\} \to 1 \text{ as } n \to \infty . \tag{3.43}$$

Let $t_k \le t \le t_{k+1}$. By (3.29),

$$(3.44)\quad r(t) - r(t_k) = \int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}\{N(t,ds) - A(t,ds) - N(t_k, ds) + A(t_k, ds)\}$$

$$+ \int_{[0,t_k]} \{\tilde{\mu}(t,s) - \tilde{\mu}(t_k,s)\}\{N(t_k, ds) - A(t_k, ds)\} .$$

By assumption, the $z_i(\cdot)$ are bounded by B and hence $\tilde{\mu}(t,s)$ and $\mu(s)$ are bounded by B. It follows that the first term is dominated by $2B\{D_{k+1} + H_{k+1}\}$ uniformly in $t_k \leq t < t_{k+1}$. By (3.41) and (3.43) it suffices to consider the second integral in (3.44). Let

$$(3.45)\quad m(t,s) = \sum_{j \in R(t,s)} e^{\beta z_j(s)} ,$$

and observe that

$$(3.46)\quad A(t,ds) = m(t,s)\, d\Lambda(s) .$$

We find from (3.9) and some algebra that uniformly in $t_k \leq t < t_{k+1}$,

$$(3.47)\quad |\tilde{\mu}(t,s) - \tilde{\mu}(t_k,s)| < 2B \left\{ \frac{m(t_{k+1}, s) - m(t_k, s)}{m(t_{k+1}, s)} \right\} .$$

Hence by (3.44) and (3.47) it suffices to show

$$(3.48)\quad P\left\{\text{For all } k \geq n, \int_{[0,t_k]} \frac{m(t_{k+1}, s) - m(t_k, s)}{m(t_{k+1}, s)} N(t_k,ds) < k^{4\varepsilon}\right\} \to 1$$

as $n \to \infty$,

and

$$(3.49)\quad P\left\{\text{For all } k \geq n,\ \int_{[0,t_k]} \frac{m(t_{k+1},\ s) - m(t_k,\ s)}{m(t_{k+1},\ s)}\, A(t_k,\ ds) < k^{4\varepsilon}\right\}$$

$$\to 1 \quad \text{as} \quad n \to \infty\ .$$

From (3.46) and some algebra we see that the k-th integral in (3.49) is majorized by H_{k+1}, so (3.49) follows from (3.43).

Now consider (3.48). It is easily seen by direct calculation that

$$L_k(s) = \int_{[0,s]} \frac{m(t_{k+1},\ u) - m(t_k,\ u)}{m(t_{k+1},\ u)}\, N(t_k,\ du) - \{N(t_{k+1},\ s) - N(t_k,\ s)\}$$

is a supermartingale for $0 \leq s \leq t_k$, which changes by jumps downward of size 1 and upward of size at most equal to 1. Furthermore, $N(t_{k+1},\ t_k) - N(t_k,\ t_k) \leq D_{k+1}$, so by (3.41), to prove (3.48) it suffices to show

$$(3.50)\qquad P\{\text{For all } k \geq n,\ L_k(t_k) < \tfrac{1}{2} k^{4\varepsilon}\} \to 1 \quad \text{as} \quad n \to \infty\ .$$

Let $S_0 = 0$, and for $j=1, 2, \ldots$ let

$$S_j = \inf\{s:\ s \geq S_{j-1},\ L_k(s) - L_k(S_{j-1}) \geq 1 \text{ or } < 0\}\ ,$$

where it is understood that $\inf \phi = t_k$. Obviously $-1 \leq L_k(S_j) - L_k(S_{j-1}) \leq 2$, and from the supermartingale property we see that on $\{S_{j-1} < t_k\}$

$$E\{L_k(S_j) - L_k(S_{j-1}) | F_{t_{k+1},\ S_{j-1}}\} \leq 0\ .$$

and hence

$$(3.51) \qquad P\{S_j < t_k,\ L_k(S_j) - L_k(S_{j-1}) \geq 1 \mid \mathcal{F}_{t_{k+1}, S_{j-1}}\} \leq \frac{1}{2} .$$

It follows from (3.51) that between downward jumps the total increase of $L_k(s)$ is stochastically less than $1+w$, where $P\{w=m\} = (\frac{1}{2})^{m+1}$, $m=0, 1, \ldots$. Since the total number of downward jumps is D_{k+1}, an easy large deviation estimate gives

$$(3.52) \qquad P\{L_k(t_k) > \frac{1}{2} k^{4\varepsilon},\ D_{k+1} \leq (k+1)^{7\varepsilon/2}\} = o\{\exp(-k^{1/7})\} .$$

Combining (3.41) and (3.52) yields (3.50), which in turn completes the proof of Proposition 3.1.

Before moving on to the other propositions, let us make a few observations about the proof just completed. The estimate in Lemma 3.8 for $E\, r^2(t)$ is very good in that it is of the same order of magnitude as $E\, r^2(t)$ itself, when there is no censoring. However, the application of the Chebyshev inequality and the addition of probabilities of "bad" sets in the proof of Lemma 3.9 is extremely crude. Only the fact that the bound of Lemma 3.8 is so much smaller than the bound in Proposition 3.1 makes it possible for such heavy-handed methods to work. The proofs of Lemmas 3.10 and 3.11 make use of very crude large deviation methods, though the sloppiness here does not cost us much more than was already lost in Lemma 3.9. The reason why such rough methods were used is that it is very difficult to say anything useful about the distrubiton of $r(t)$ as a stochastic process in t. However, it seems clear that the $r(t)$ process is

actually much smaller than is claimed by Proposition 3.1. In fact, it is probably $o\{D^{\varepsilon}(t)\}$ for any $\varepsilon > 0$. This suggests that the error committed by approximating $\dot{\ell}(t,\beta_0)$ by a martingale may be quite small even when the sample sizes are moderate.

3.3. Approximation of $\langle Q \rangle(t)$ by $-\ddot{\ell}(t,\beta)$

This section will prove Propositions 3.2 and 3.3. Proposition 3.2 shows that $\langle Q \rangle(t) + \ddot{\ell}(t,\beta)$ is $o\{D(t)^{1-2\varepsilon}\}$. The proof of Proposition 3.2 is made relatively painless by the observation that the techniques used in the proof of Proposition 3.1 apply almost without change. I am sure that the reader who has come this far will be grateful for not having to go through another proof as technical and complicated as that of Proposition 3.1.

Proposition 3.2. As $K \to \infty$

$$P\{|\ddot{\ell}(t,\beta) + \langle Q \rangle(t)| \leq K + D(t)^{1-2\varepsilon},\ t \geq 0\} \to 1 ,$$

uniformly in B-experiments.

Proof. Note from (3.10) and (3.28) that

$$-\ddot{\ell}(t,\beta) = \int_{[0,t]} \tilde{\sigma}^2(t,s)\, N(t,ds) \tag{3.53}$$

where

$$\tilde{\sigma}^2(t,s) = \frac{\sum_{j \in R(t,s)} z_j^2(s)\, e^{\beta z_j(s)}}{\sum_{j \in R(t,s)} e^{\beta z_j(s)}} - \tilde{\mu}^2(t,s) . \tag{3.11}$$

Now apply some algebra to (3.30).

$$(3.30)\quad \langle Q\rangle(t) = \sum_i \int_{[0,t]} \{z_i(s) - \mu(s)\}^2 \, A_i(t,ds)$$

$$= \sum_i \int_{[0,t]} \{z_i(s) - \tilde{\mu}(t,s) + \tilde{\mu}(t,s) - \mu(s)\}^2 \, A_i(t,ds)$$

$$= \int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}^2 \, A(t,ds)$$

$$+ \sum_i \int_{[0,t]} \{z_i(s) - \tilde{\mu}(t,s)\}^2 \, A_i(t,ds)$$

$$= \int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}^2 \, A(t,ds)$$

$$+ \int_{[0,t]} \tilde{\sigma}^2(t,s) \, A(t,ds) \, .$$

It follows from some additional algebra that

$$(3.54)\quad \ddot{\ell}(t,\beta) + \langle Q\rangle(t) = \int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}^2 \, A(t,ds)$$

$$+ \int_{[0,t]} \{\tilde{\mu}^2(t,s) - \mu^2(s)\}\{N(t,ds) - A(t,ds)\}$$

$$+ \int_{[0,t]} \left\{ \mu_2(s) - \frac{\sum_{j\in R(t,s)} z_j^2(s)\, e^{\beta z_j(s)}}{\sum_{j\in R(t,s)} e^{\beta z_j(s)}} \right\} \{N(t,ds) - A(t,ds)\}$$

$$- \int_{[0,t]} \sigma^2(s)\{N(t,ds) - A(t,ds)\} \, ,$$

where

$$\mu_2(s) = \frac{E\{z_1^2(s)\, e^{\beta z_1(s)} \; ; \; x_1 \wedge c_1 \geq s\}}{E\{e^{\beta z_1(s)} \; ; \; x_1 \wedge c_1 \geq s\}}$$

and

$$\sigma^2(s) = \mu_2(s) - \mu^2(s) \; .$$

The first three terms on the right-hand side of (3.54) can be estimated by techniques similar to the proof of Proposition 3.1. The first term has the same expectation as

$$\int_{[0,t]} \{\tilde{\mu}(t,s) - \mu(s)\}^2 \, N(t,ds) \; ,$$

and the expectation of this was shown in the proof of Lemma 3.8 to be bounded above by const.$\{3 + \log D(t)\}$. If we use the same $\{t_k\}$ sequence as was defined in (3.34), then the Markov inequality implies

$$(3.55) \qquad P[\int_{[0,t_k]} \{\tilde{\mu}(t,s) - \mu(s)\}^2 \, A(t_k,ds) > \tfrac{1}{2} D^{1-2\varepsilon}(t_k)]$$

$$\leq \text{const.} \; \frac{3 + \log D(t_k)}{D(t_k)^{1-2\varepsilon}}$$

$$\leq \text{const.} \; \frac{3 + (1+3\varepsilon)\log k}{k^{1+\varepsilon-6\varepsilon^2}}$$

for $t_k < \infty$. The rest follows as before.

The second term follows directly from the techniques of Proposition 3.1. The integrand has the form a^2-b^2, and at certain points in

the proof one uses the factorization $a^2-b^2 = (a+b)(a-b)$ and the boundedness of $(a+b)$. Bounding the third term requires almost no modifications of the proof of Proposition 3.1.

The last term is a martingale in t, and the Kolmogorov inequality yields an easy proof that this term is $o\{D^{\frac{1}{2}+\epsilon}(t)\}$.

In (3.54), the term for which our upper bound was the largest was the first term. However, the argument that was given at the end of Section 3.2 for why $r(t)$ is actually much smaller than $D^{\frac{1}{2}-\epsilon}(t)$ also applies here, so that this first term is probably smaller than $D^{\frac{1}{2}}(t)$ and possibly much smaller. The proof of Proposition 3.1 shows that the second and third terms of (3.54) are smaller than $D^{\frac{1}{2}-\epsilon}(t)$, and it seems clear that these terms are actually much smaller and therefore quite negligible. The fourth term, however, is in fact larger than $o\{D^{\frac{1}{2}}(t)\}$. Thus, Proposition 3.2 probably holds if we replace $D^{1-2\epsilon}(t)$ by $D^{\frac{1}{2}+\epsilon}(t)$, but it does not hold with $D^{\frac{1}{2}}(t)$.

In order to attain our goal of approximating $\dot{\ell}(t,\beta)$ by $W\{-\ddot{\ell}(t,\beta)\}$, where $W(\cdot)$ is a standard Brownian motion, we need to know that

$$r(t) = \dot{\ell}(t,\beta) - Q(t)$$

and

$$\ddot{\ell}(t,\beta) + \langle Q \rangle(t)$$

are sufficiently small relative to $\langle Q \rangle(t)$, whereas what we have from Propositions 3.1 and 3.2 is that these processes are small relative to $D(t)$. Proposition 3.3 shows that the $\langle Q \rangle(t)$ process is of the same order of magnitude as the function $D(t)$, and this implies what

we need. The rather technical assumption made in (3.3) and (3.4) about $\text{var}\{z_1(s) | x_1 \wedge c_1 \geq s\}$ for small s is necessary in order to guarantee that Proposition 3.3 holds uniformly in configurations of entry times. It seems probable that Propositions 3.1 and 3.2 hold uniformly in B-experiments with $D(t)$ replaced by $\langle Q \rangle(t)$ or $-\ddot{\ell}(t,\beta)$, and in this case, assumptions (3.3) and (3.4) and Proposition 3.3 would all be unnecessary. However, such a change in Proposition 3.1 would have introduced considerable complications into an already complicated proof. In any case, all but the most suspicious or masochistic readers are encouraged to skip the rather tedious proof of Proposition 3.3 and move on to the more exciting and enlightening Section 3.4.

Proposition 3.3. There exists $\alpha > 0$ such that, as $K \to \infty$,

$$P\{\langle Q \rangle(t) + K > \alpha D(t), \ t \geq 0\} \to 1 \ ,$$

uniformly in (B,δ,η)-experiments.

Proof. Assume that $D(\infty) = \infty$ and define t_n, $n=1, 2, 3, \ldots$, to satisfy

$$D(t_n) = 2^n \ . \tag{3.56}$$

Then by the definition (3.27) of $D(t)$,

(3.57) $$\sum_i E\, N_i(t_n,t_n) = 2^n .$$

Let

(3.58) $$\gamma = 1 - \exp\{-e^{-|\beta|B}\, \Lambda(\delta)\} ,$$

and note that $\gamma > 0$ by (3.3). It will be necessary to show that

(3.59) $$E\, N_i(t,\delta) \geq \gamma\, E\, N_i(t,t) \quad \text{for all} \quad t \geq 0 .$$

To do this, recall from (3.5) that

(3.60) $$N(t,s) = I_{\{x_i \leq s \wedge c_i \wedge (t-y_i)^+\}} .$$

Condition on c_i and $(t-y_i)^+$. If $\delta \geq c_i \wedge (t-y_i)^+$, then $N_i(t,\delta) = N_i(t,t)$. Otherwise, $N_i(t,\delta) = I_{\{x_i \leq \delta\}}$ and

(3.61)
$$\begin{aligned} E\, N_i(t,\delta) &= P\{x_i < \delta\} \\ (\text{by (3.2)}) \quad &= P\{\Lambda_i(\delta) \leq h_i\} \\ &\geq P\{e^{-|\beta|B}\, \Lambda(\delta) \leq h_i\} \\ &\geq \gamma \\ &\geq \gamma\, E\, N_i(t,t) , \end{aligned}$$

so that (3.59) follows easily.

It follows from (3.57) and (3.59) that

(3.62) $$\sum_i E\, N_i(t_n,\delta) \geq \gamma\, 2^n ,$$

and by the martingale property of (3.18) we have

$$\sum_i E\, A_i(t_n,\delta) \geq \gamma\, 2^n \,. \tag{3.63}$$

By (3.4),

$$P\{|z_i(s)-\mu(s)| > \frac{\eta}{2B+2} \,\Big|\, x_i \wedge c_i \geq s\} \geq \frac{\eta^2}{8+8B^2} \quad \text{for} \quad 0 \leq s \leq \delta \,. \tag{3.64}$$

Thus, we can find $\rho > 0$ such that

$$P\{|z_i(s)-\mu(s)| > \rho \,\Big|\, x_i \wedge c_i \geq s\} \geq \rho \quad \text{for} \quad 0 \leq s \leq \delta \,. \tag{3.65}$$

Again using the martingale property of (3.18), we get

$$E \sum_i \int_0^\delta I_{\{|z_i(s)-\mu(s)|>\rho\}}\, N_i(t_n,ds) \tag{3.66}$$

$$= E \sum_i \int_0^\delta I_{\{|z_i(s)-\mu(s)|>\rho\}}\, A_i(t_n,ds)$$

$$\geq e^{-|\beta|B} \sum_i E \int_0^{\delta\wedge(t_n-y_i)^+} I_{\{|z_i(s)-\mu(s)|\geq\rho\}}\, I_{\{x_i\wedge c_i\geq s\}}\, \Lambda(ds)$$

$$\text{(by Fubini)} \geq e^{-|\beta|B} \sum_i \int_0^{\delta\wedge(t_n-y_i)^+} P\{|z_i(s)-\mu(s)| > \rho \,\Big|\, x_i \wedge c_i \geq s\}$$

$$P\{x_i \wedge c_i \geq s\}\, \Lambda(ds)$$

$$\text{(by (3.65))} \geq \rho e^{-|\beta|B} \sum_i \int_0^{\delta\wedge(t-y_i)^+} P\{x_i \wedge c_i \geq s\}\, \Lambda(ds)$$

$$\geq \rho e^{-2|\beta|B} \sum_i E \int_0^\delta A_i(t_n,ds)$$

$$\geq \rho e^{-2|\beta|B} \sum_i E\, A_i(t_n,\delta)$$

(by (3.63)) $\geq \rho e^{-2|\beta|B}\, \gamma\, 2^n$.

The random variable

$$V_n = \sum_i \int_0^\delta I_{\{|z_i(s)-\mu(s)|>\rho\}}\, N_i(t_n,ds)$$

appearing on the left-hand side of (3.66) is a sum of independent Bernoulli variables, so that its variance is less than its expectation. It is easy to show, by the Chebyshev inequality, that V_n exceeds $\frac{1}{2} E\, V_n$ except for finitely many n, almost surely. But $\rho^2 V_n$ is dominated by

$$\sum_i \int_0^{t_n} \{z_i(s) - \mu(s)\}^2\, N_i(t_n,ds)$$

so that, except for finitely many n,

$$(3.67) \quad \sum_i \int_0^{t_n} \{z_i(s) - \mu(s)\}^2\, N_i(t_n,ds) > \tfrac{1}{2} \rho^3\, e^{-2|\beta|B}\, \gamma\, 2^n .$$

Using Kolmogorov's inequality and the fact that

$$\sum_i \int_0^t \{z_i(s) - \mu(s)\}^2\, \{N_i(t,ds) - A_i(t,ds)\}$$

is a martingale in t shows that, except for finitely many n,

$$(3.68) \quad \sum_i \int_0^{t_n} \{z_i(s) - \mu(s)\}^2\, A_i(t_n,ds) \geq \tfrac{1}{4} \rho^3\, e^{-2|\beta|B}\, \gamma\, 2^n .$$

Since the left side of (3.68) is $\langle Q \rangle(t_n)$ and the right side is const. $D(t_n)$, both of which are increasing in n, the Proposition follows easily. Minor changes take care of the $D(\infty) < \infty$ case.

3.4. Consistency of $\hat{\beta}(t)$ and Approximation of $-\ddot{\ell}\{t,\hat{\beta}(t)\}\{\hat{\beta}(t) - \beta\}$ by $\dot{\ell}(t,\beta)$

The proofs of Propositions 3.4 and 3.5 make use of the following lemma concerning the third derivative of the log partial likelihood.

Lemma 3.12.

$$|\dddot{\ell}(t,\beta)| \leq 2B\{-\ddot{\ell}(t,\beta)\} ,$$

where this holds for all values of β, not just the true value.

Proof. By some algebra

$$\dddot{\ell}(t,\beta) = \sum_i \int_{[0,t]} \tilde{\mu}_3(t,s)\, N_i(t,ds) , \tag{3.69}$$

where

$$\tilde{\mu}_3(t,s) = \frac{\sum_{j \in R(t,s)} \{z_j(s) - \tilde{\mu}(t,s)\}^3 e^{\beta z_j(s)}}{\sum_{j \in R(t,s)} e^{\beta z_j(s)}} . \tag{3.70}$$

The fact that

$$|z_j(s) - \tilde{\mu}(t,s)| \leq 2B$$

implies

$$|\tilde{\mu}_3(t,s)| \leq 2B\, \tilde{\sigma}^2(t,s) , \tag{3.71}$$

where $\tilde{\sigma}^2(t,s)$ is given by (3.11). The lemma now follows from (3.10).

Proposition 3.4. As $L \to \infty$

$$P[|\hat{\beta}(t) - \beta_0| < \{-\ddot{\ell}(t,\beta)\}^{\varepsilon-\frac{1}{2}} \text{ for } -\ddot{\ell}(t,\beta) > L] \to 1 ,$$

uniformly in (B,δ,η)-experiments.

Proof. The proof is based on the following observation. Since $-\ddot{\ell}(t, \beta+\Delta)$ is nonnegative for all $t \geq 0$ and $\Delta \in \mathbb{R}$, $\dot{\ell}(t, \beta+\Delta)$ is nonincreasing and continuous in Δ. Thus, if for some $\delta > 0$ we have

$$\dot{\ell}(t, \beta-\delta) > 0 > \dot{\ell}(t, \beta+\delta) , \tag{3.72}$$

then $\hat{\beta}(t)$ exists and

$$|\hat{\beta}(t) - \beta| < \delta .$$

Furthermore, if (3.72) holds, then $\dot{\ell}(t, \beta+\Delta)$ is strictly decreasing and $\hat{\beta}(t)$ is unique. The proof proceeds by showing that $\dot{\ell}(t,\beta)$ is small compared to $-\ddot{\ell}(t,\beta)$, so that (3.72) holds for a small δ.

An easy argument using Kolmogorov's inequality shows

$$P[|Q(t)| \leq \tfrac{1}{2}\{\langle Q\rangle(t)\}^{(\varepsilon+1)/2} \text{ for } \langle Q\rangle(t) \geq L] \to 1 \text{ as } L \to \infty . \tag{3.73}$$

By Propositions 3.1, 3.2, and 3.3 we can replace $Q(t)$ by $\dot{\ell}(t,\beta)$ and $\langle Q\rangle(t)$ by $-\ddot{\ell}(t,\beta)$ in (3.73) to get

$$P[|\dot{\ell}(t,\beta)| \leq \{-\ddot{\ell}(t,\beta)\}^{(\varepsilon+1)/2} \text{ for } \ddot{\ell}(t,\beta) \geq L] \to 1 , \tag{3.74}$$

uniformly in (B,δ,η)-experiments.

For $\delta > 0$, there exists δ^* with $\delta > \delta^* > 0$ such that

$$\dot{\ell}(t, \beta+\delta) = \dot{\ell}(t,\beta) + \delta\, \ddot{\ell}(t, \beta+\delta^*) \ . \tag{3.75}$$

From Lemma 3.12, it is easy to show that

$$-\ddot{\ell}(t, \beta+\delta^*) > -\ddot{\ell}(t,\beta)\ (1 - 2B\delta) \ . \tag{3.76}$$

By (3.74) and (3.76),

$$\dot{\ell}(t, \beta+\delta) \leq \{-\ddot{\ell}(t,\beta)\}^{(\varepsilon+1)/2} - \delta\{-\ddot{\ell}(t,\beta)\}\ (1 - 2B\delta) \tag{3.77}$$

holds with high probability for all $\delta > 0$ and for all t satisfying $-\ddot{\ell}(t,\beta) > L$ when L is sufficiently large. If we set $\delta = \{-\ddot{\ell}(t,\beta)\}^{\varepsilon-\frac{1}{2}}$ in (3.77), then the right side becomes

$$\{-\ddot{\ell}(t,\beta)\}^{(\varepsilon+1)/2}\ [1 - \{-\ddot{\ell}(t,\beta)\}^{\varepsilon/2}\ \{1 - 2B(-\ddot{\ell}(t,\beta))^{\varepsilon-\frac{1}{2}}\}] \ . \tag{3.78}$$

If L is sufficiently large, then (3.78) is negative whenever $-\ddot{\ell}(t,\beta) \geq L$. This, together with a similar argument for $\dot{\ell}(t, \beta-\delta)$, shows that

$$\dot{\ell}[t,\beta - \{-\ddot{\ell}(t,\beta)\}^{\varepsilon-\frac{1}{2}}] > 0 > \dot{\ell}[t,\beta + \{-\ddot{\ell}(t,\beta)\}^{\varepsilon-\frac{1}{2}}] \tag{3.79}$$

holds with high probability for all t satisfying $-\ddot{\ell}(t,\beta) > L$ when L is sufficiently large. The remarks at the beginning of the proof show that this suffices.

Proposition 3.5. As $L \to \infty$

$$P[|\ddot{\ell}\{t, \hat{\beta}(t)\}\{\hat{\beta}(t) - \beta\} + \dot{\ell}(t,\beta)| < \{-\ddot{\ell}(t,\beta)\}^{3\epsilon} \text{ for } -\ddot{\ell}(t,\beta) > L] \to 1 ,$$

uniformly in (B,δ,η)-experiments.

Proof. Note that

$$0 = \dot{\ell}\{t, \hat{\beta}(t)\} = \dot{\ell}(t,\beta) + \ddot{\ell}(t, \beta+\delta) \{\hat{\beta}(t) - \beta\} , \tag{3.80}$$

where $\beta+\delta$ is between β and $\hat{\beta}(t)$. Thus, the quantity in absolute value signs in Proposition 3.5 equals

$$[\ddot{\ell}\{t, \hat{\beta}(t)\} - \ddot{\ell}(t, \beta+\delta)] \{\hat{\beta}(t) - \beta\} . \tag{3.81}$$

By Lemma 3.12, the absolute value of (3.81) is less than

$$2B\{-\ddot{\ell}(t, \beta+\delta')\} \{\hat{\beta}(t) - \beta\}^2 , \tag{3.82}$$

where $\beta+\delta'$ is again between β and $\hat{\beta}(t)$. Another application of Lemma 3.12 shows

$$-\ddot{\ell}(t, \beta+\delta') < e^{2B|\delta'|} \{-\ddot{\ell}(t,\beta)\} , \tag{3.83}$$

so that (3.82) is less than

$$2B\, e^{2B|\delta'|} \{-\ddot{\ell}(t,\beta)\} \{\hat{\beta}(t) - \beta\}^2 . \tag{3.94}$$

Proposition 3.5 now follows from Proposition 3.4.

We will also need Lemma 3.13 when we prove Theorem 3.7.

Lemma 3.13. As $K \to \infty$,

$$P[|\ddot{\ell}\{t, \hat{\beta}(t)\} - \ddot{\ell}(t,\beta)\} \leq K + \{-\ddot{\ell}(t,\beta)\}^{(1+2\varepsilon)/2}, \ t \geq 0] \to 1 ,$$

uniformly in (B,δ,η)-experiments.

Proof. An argument like the proof of Proposition 3.5 shows that as $L \to \infty$,

$$P[|\ddot{\ell}\{t, \hat{\beta}(t)\} - \ddot{\ell}(t,\beta)| \leq \{-\ddot{\ell}(t,\beta)\}^{(1+2\varepsilon)/2} \quad \text{for } -\ddot{\ell}(t,\beta) > L] \to 1 , \tag{3.85}$$

uniformly in (B,δ,η)-experiments. But

$$|\ddot{\ell}\{t, \hat{\beta}(t)\} - \ddot{\ell}(t,\beta)| \leq 2B^2 N(t) \tag{3.86}$$

and it is easy to use the Chebyshev inequality to show that, as $K \to \infty$,

$$P\{N(t) \leq K + 2D(t), \ t \geq 0\} \to 1 . \tag{3.87}$$

By combining Propositions 3.2 and 3.3, we get

$$P\{-\ddot{\ell}(t,\beta) + K > \frac{\alpha}{2} D(t), \ t \geq 0\} \to 1 \tag{3.88}$$

as $K \to \infty$. Combining (3.86), (3.87), and (3.88) yields

$$P[|\ddot{\ell}\{t, \hat{\beta}(t)\} - \ddot{\ell}(t,\beta)| \leq K + \frac{8B^2}{\alpha} \{-\ddot{\ell}(t,\beta)\}, \ t \geq 0] \to 1 , \tag{3.89}$$

as $K \to \infty$. Together, (3.85) and (3.89) imply the lemma.

3.5. Approximation of the Martingale Q(t) by a Brownian Motion

It remains to introduce the Brownian motion whose existence is

claimed by Proposition 3.6. The most natural way of doing this would be to use Theorem A.3 to embed $Q(t)$ in a Brownian motion $W(\cdot)$. The result would be a continuous time analog of Proposition 2.1. Proposition 3.6 would then follow from a proof almost exactly like that of Theorem 2.2. Unfortunately, I don't know how to prove Theorem A.3, even though my intuition tells me it must be true. Therefore, in order to placate those narrowminded readers who refuse to trust my intuition, I shall give a somewhat messier proof of Proposition 3.6 based on Theorem A.2, for which I purport to have a proof in the Appendix.

Proposition 3.6. On an enlarged version of our probability space, there exists a standard Brownian motion $W(\cdot)$ such that, as $K \to \infty$,

$$P\{|Q(t) - W\{\langle Q\rangle(t)\}| < K + \langle Q\rangle(t)^{\frac{1}{4}+\varepsilon}, \text{ for } t \geq 0\} \to 1 .$$

This convergence holds uniformly in B-experiments.

Proof. The strategy will be as follows. First, a discretized version of $Q(t)$ will be defined. The predictable quadratic variation of this discrete time martingale is shown to be close to $\langle Q\rangle(t)$, the predictable quadratic variation process of $Q(t)$. Theorem A.2 is then used to embed the discrete time martingale in a Brownian motion. Finally, the proof of Theorem 2.2 is used to finish the proof of Proposition 3.6.

Define $\mathcal{F}_t$ stopping times $0 = t_0 < t_1 < \ldots$ by

$$(3.90)\quad t_{k+1} = \inf[t: t>t_k, \{\langle Q\rangle(t) - \langle Q\rangle(t_k)\} \wedge |Q(t) - Q(t_k)| \geq 1] .$$

Let $X_k = Q(t_k) - Q(t_{k-1})$, $F_k^d = F_{t_k}$, and $v_k = E(X_k^2 | F_{k-1}^d)$.

Note that $Q^2(t_k) - \langle Q \rangle(t_k)$ is an F_k^d martingale, as are

$$Q^2(t_k) - \sum_{i=1}^{k} X_i^2 \quad \text{and} \quad \sum_{i=1}^{k} X_i^2 - \sum_{i=1}^{k} v_i .$$

It follows that

$$\langle Q \rangle(t_k) - \sum_{i=1}^{k} v_i$$

is an F_k^d martingale. Define

$$\Delta\langle Q \rangle_k = \langle Q \rangle(t_k) - \langle Q \rangle(t_{k-1}) .$$

Then since $\Delta Q_k \leq 1$,

$$E\{(\Delta\langle Q \rangle_k - v_k)^2 | F_{k-1}^d\} \leq E\{(\Delta\langle Q \rangle_k)^2 | F_{k-1}\} \tag{3.91}$$

$$\leq E\{\Delta\langle Q \rangle_k | F_{k-1}\} = v_k .$$

An application of Kolmogorov's inequality now shows that

$$P\{|\langle Q \rangle(t_k) - \sum_{i=1}^{k} v_i| \leq K + (\sum_{i=1}^{k} v_i)^{\frac{1}{2}+\varepsilon} \text{ for all } k \geq 0\} \to 1 \tag{3.92}$$

as $K \to \infty$, uniformly in B-experiments.

It follows immediately from Theorem A.2 that there exists a standard Brownian motion $W(\cdot)$ and a sequence of random variables $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \ldots$ on an enlarged version of our probability space such that (3.93) holds. (I won't bother to put stars on everything this time.)

(3.93a) $$X_k = W(\tau_k) - W(\tau_{k-1})$$

(3.93b) $$E(\tau_k - \tau_{k-1} | F^d_{k-1}) = v_k$$

(3.93c) $$\mathrm{var}(\tau_k - \tau_{k-1} | F^d_{k-1}) \le 2(B+1)^2 v_k$$

(3.93d) τ_k is F^d_k measurable and the pre-τ_k σ-algebra of $W(\cdot)$ is contained in F^d_k.

The situation is now very similar to that of Theorem 2.2. We know that

(3.94) $$Q(t_k) = W(\tau_k) ,$$

and it remains to be shown that

(3.95) $$\tau_k - \langle Q \rangle(t_k)$$

is sufficiently small so that $Q(t) - W\{\langle Q \rangle(t)\}$ is small. Note that (3.95) is an F^d_k-martingale, and that by (3.91) and (3.93c)

(3.96) $$\mathrm{var}\left[\{\tau_k - \langle Q \rangle(t_k)\} - \{\tau_{k-1} - \langle Q \rangle(t_{k-1})\} \middle| F^d_{k-1}\right] \le 4(B+1)^2 v_k + 2v_k .$$

Define $k(u)$ as in (2.15):

(3.97) $$k(u) = \sup\{k: \sum_{i=1}^{k} v_i \le u\} .$$

By (3.96) and Kolmogorov's inequality, we get the following analog of (2.22) for $m=1, 2, \ldots$.

$$(3.98)\quad P\{\sup_{k\le k(2^m)} |\tau_k - \langle Q\rangle(t_k)| \ge 2^{m(1+\varepsilon)/2}\} \le 2^{\frac{1}{m\varepsilon}}\{4(B+1)^2 + 2\}\ .$$

The rest of the proof is as in Theorem 2.2, except that at the end one has to refer to (3.92) to justify replacing $(\sum_{i=1}^{k} v_i)^{\frac{1}{4}+\varepsilon}$ by $\langle Q\rangle(t)^{\frac{1}{4}+\varepsilon}$.

3.6. The Main Theorem

Theorem 3.7. Let $0 < \varepsilon < 15$. On an enlarged version of our probability space, there exists a standard Brownian motion $W(\cdot)$ such that, as $K\to\infty$,

$$P[|\dot{\ell}(t,\beta) - W\{-\ddot{\ell}(t,\beta)\}| \le K + \{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}-\varepsilon} \text{ for all } t \ge 0] \to 1$$

and, as $L\to\infty$,

$$P\{|[-\ddot{\ell}\{t, \hat{\beta}(t)\}] \cdot \{\hat{\beta}(t) - \beta\} - W[-\ddot{\ell}\{t, \hat{\beta}(t)\}]|$$

$$\le [-\ddot{\ell}\{t, \hat{\beta}(t)\}]^{\frac{1}{2}-\varepsilon} \text{ for } -\ddot{\ell}\{t, \hat{\beta}(t)\} > L] \to 1\ .$$

The convergence is uniform in (B,δ,η)-experiments, and for β in compact intervals.

Proof. Let $0 < \varepsilon < \varepsilon' < \varepsilon'' < \varepsilon''' < \frac{1}{15}$. Since Propositions 3.1 and 3.2 hold with ε replaced by ε''', Propositions 3.1, 3.2, and 3.3 imply

$$(3.99)\quad P[|r(t)| \le K + \{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}-\varepsilon''},\ t \ge 0] \to 1$$

and

$$(3.100) \qquad P[|\ddot{\ell}(t,\beta) + \langle Q\rangle(t)| \leq K + \{-\ddot{\ell}(t,\beta)\}^{1-2\varepsilon''}] \to 1$$

as $K\to\infty$, where the convergence is of course uniform in (B,δ,η)-experiments. A proof like that of Theorem 2.2 together with (3.100) implies that as $K\to\infty$,

$$(3.101) \quad P[|W\{\langle Q\rangle(t)\} - W\{-\ddot{\ell}(t,\beta)\}| < K + \{-\ddot{\ell}(t,\beta)\}^{\frac{1}{2}-\varepsilon'},\ t \geq 0] \to 1 .$$

The first part of Theorem 3.7 now follows from (3.99), (3.100), (3.101), and Proposition 3.6.

The second part of Theorem 3.7 follows easily from Proposition 3.5, Lemma 3.13 at the end of Section 3.4, and the first part of the theorem.

The operational conclusions to be drawn from Theorem 3.7 are the same as those which were drawn for the case of simultaneous entry at the end of Chapter II. Again, the behavior of

$$\dot{\ell}(t,\beta)$$

and of

$$[-\ddot{\ell}\{t,\ \hat{\beta}(t)\}]\ \{\hat{\beta}(t) - \beta\}$$

in information time is approximately that of a standard Brownian motion until a possibly infinite stopping time. The problem of sequentially estimating β or testing $H_0\colon \beta = \beta_0$ is asymptotically equivalent to the problem of estimating or testing the drift of a possibly stopped Brownian motion.

The sharp-eyed reader will have noticed that the exponent $\frac{1}{4} + \varepsilon$ found in (2.25) and in Theorem 2.3 of Chapter II has been replaced by $\frac{1}{2} - \varepsilon$ in Theorem 3.7. This difference is due to the large exponents of $D(t)$ which are present in Propositions 3.1 and 3.2. However, it was argued in Sections 3.2 and 3.4 that $\dot{\ell}(t,\beta)$ and $-\ddot{\ell}(t,\beta)$ are in fact much more closely approximated by $Q(t)$ and $\langle Q \rangle(t)$ than is claimed in Propositions 3.1 and 3.2. Thus, it is possible that Theorem 3.7 remains true when the exponent $\frac{1}{2} - \varepsilon$ is replaced by $\frac{1}{4} + \varepsilon$, and the Brownian motion approximation may be almost as good in the case of staggered entry as it is in the case of simultaneous entry. It seems desirable to conduct a Monte Carlo experiment to get some feeling for the practical limitations of Theorem 3.7. For the problem of testing the null hypothesis $\beta=0$, Gail, DeMets, and Slud (1981) conclude that the score statistic under the null hypothesis is reasonably approximated by a Brownian motion. Their time renormalization is not appropriate for general β, however.

3.7. Multidimensional Covariates

It is easy to generalize the notation and formulation of the model given in Section 3.1 to the case of p-dimensional β and $z_i(\cdot)$, $p \geq 2$. The only nonobvious change is in the technical assumption given previously in formulas (3.3) and (3.4). For p-dimensional covariate processes $z_i(\cdot)$ we will assume that there exist positive numbers δ and η such that

$$\Lambda(\delta) > 0 \tag{3.102}$$

and

$$(3.103)\quad \|\mathrm{cov}\{z_1(s) \mid x_1 \wedge c_1 \geq s\}\|_{\min} \geq \eta^2 , \quad \text{for } 0 \leq s \leq \delta ,$$

where $\|M\|_{\min}$ denotes the minimum eigenvalue of the matrix M. The p-dimensional martingale Q(t) and its p×p matrix-valued predictable quadratic covariation process $\langle Q \rangle(t)$ are easy generalizations of the one-dimensional versions. It also seems that Propositions 3.1 through 3.5 go through essentially as before. Difficulty is encountered when one attempts to generalize Proposition 3.6. The author is not aware of any techniques for embedding a p-dimensional martingale in a p-dimensional Brownian motion.

Cox (1963) and Whitehead (1978) discuss large sampel sequential tests of hypotheses in the presence of nuisance parameters, and it seems natural to adopt a similar approach here. Suppose that the first coordinate of the p-vector β is of primary interest, with the other coordinates being regarded as nuisance parameters. This would typically be the case when the first coordinate is a treatment indicator in a clinical trial comparing two treatments. Let us adopt the notation of Cox (1963) and Whitehead (1978) and use $\theta \in \mathbb{R}$ to refer to the first coordinate of β and $\phi = (\phi_1, \ldots, \phi_{p-1})$ to refer to the other coordinates. Let $\ell(t,\theta,\phi)$ be the log partial likelihood at time t. The derivatives of $\ell(t,\theta,\phi)$ are written in the usual way as $\ell_\theta(t,\theta,\phi)$, $\ell_\phi(t,\theta,\phi)$, $\ell_{\theta\theta}(t,\theta,\phi)$, etc. Let $1/\ell^{\theta\theta}(t,\theta,\phi)$ be the leading element of the inverse of the matrix of second derivatives, so that

$$\ell^{\theta\theta} = \ell_{\theta\theta} - (\ell_{\theta\phi})^T (\ell_{\phi\phi})^{-1} (\ell_{\theta\phi}) . \tag{3.104}$$

(cf. Whitehead (1978), p. 352). Recall that in the one-dimensional case, our information time $-\ddot{\ell}(t,\beta)$ was approximately equal to the reciprocal of the variance of $\hat{\beta}(t)$. Here, the variance of $\hat{\theta}(t)$, the maximum partial likelihood estimator of θ, is approximately $-1/\ell^{\theta\theta}(t,\theta,\phi)$, so that the natural approach is to use $-\ell^{\theta\theta}(t,\theta,\phi)$ as an information time. Define, for $u \geq 0$,

$$\tau(u) = \inf\{t: -\ell^{\theta\theta}(t,\hat{\theta},\hat{\phi}) \geq u\} . \tag{3.105}$$

Assume that there are infinitely many entry times, so that $\tau(u)$ is finite for all $u \geq 0$.

Preliminary Theorem 3.14. Suppose that we have a sequence of (B,δ,η)-experiments indexed by n. Fix $u_* > 0$. Then as $n \to \infty$,

$$(\cdot)\, n^{\frac{1}{2}}[\hat{\theta}\{\tau((\cdot)n)\} - \theta] \xrightarrow{\mathcal{L}} W(\cdot) \tag{3.106}$$

on $[u_*,\infty)$, where $W(\cdot)$ is standard Brownian motion.

It is not hard to show that the convergence of (3.106) holds for a single, fixed u, but the remainder of the proof has not yet been worked out.

APPENDIX

A.1. Basic Facts About Martingales

Chapter 2 of Gill (1980) contains a nice summary of the facts about martingales and stochastic integrals used in this dissertation. A more detailed development is found in Liptser and Shiryayev (1977) and Liptser and Shiryayev (1978). However, for the convenience of the reader, we will include a very brief review, most of which is paraphrased from Gill (1980).

Let (Ω,F,P) be a complete probability space. Let $\{F_t, t \in [0,\infty)\}$ be a right-continuous, increasing family of σ-algebras. Generally, F_t is thought of as being generated by events occurring in the time interval $[0,t]$. A stochastic process $X(t) = X(t,\omega)$, $t \in [0, \infty)$, is said to be adapted to $\{F_t\}$ if $X(t)$ is F_t-measurable for each t. A process $Y(t,\omega)$ is defined by Gill (1980) to be F_t-predictable if, as a function on $[0,\infty) \times \Omega$, it is measurable with respect to the σ-algebra generated on $[0,\infty) \times \Omega$ by all adapted processes with left-continuous paths. An adapted process $M(t)$ is an F_t-martingale if it is right-continuous with left-hand limits and satisfies

$$E\{M(t)|F_s\} = M(s) , \quad s \leq t .$$

A martingale M is said to be square-integrable if

$$\sup_{t[0,\infty)} E\{M^2(t)\} < \infty .$$

If M is a square-integrable martingale, it follows easily from Jensen's inequality that M^2 satisfies

$$E\{M^2(t)|F_s\} \geq M^2(s) , \quad s \leq t ,$$

so that M^2 is a submartingale. By the Doob-Meyer decomposition theorem, (see Lipster and Shiryayev (1977), Corollary and Note on page 68) there exists a predictable, increasing process $\langle M \rangle(t)$ such that

$$M^2(t) - \langle M \rangle(t)$$

is a martingale. The process $\langle M \rangle(t)$ will be referred to in this dissertation as the predictable quadratic variation process of $M(t)$.

Let $X(t)$ and $Y(t)$ be stochastic processes with piecewise continuous paths. Suppose further that the paths of $X(t)$ are of bounded variation on bounded intervals. Then one can define the process

$$U(t) = \int_{s\in[0,t]} Y(s)\, dX(s)$$

by taking Stieltjes integrals pathwise. If $X(t)$ is a square-integrable martingale, and $Y(t)$ is bounded and predictable, then $U(t)$ is itself a square-integrable martingale with predictable quadratic variation process

$$\langle U \rangle(t) = \int_{s\in[0,t]} Y^2(s)\, d\langle X \rangle(s) .$$

According to Gill (1980), p. 9, "a multivariate counting process $N = \{N_i : i=1, \ldots, r\}$ is a finite family of adapted processes N_i such that for almost all $\omega \in \Omega$, the paths of $N_1, \ldots, N_r$ are

nondecreasing, right continuous, integer-valued functions, zero at time zero, and with jumps of size +1 only, no two processes jumping at the same time." Furthermore, "there exist right continuous, nondecreasing, predictable processes A_i, zero at a time zero, such that

$$M_i = N_i - A_i , \quad i=1, \ldots, r$$

are local martingales." (Gill (1980), p. 12). These definitions are satisfied by the N_i's and A_i's defined on the bottom of page 40 in Chapter III. Since in our case the M_i's are locally square-integrable and the A_i's are continuous, it follows from Theorem 2.3.1 on page 12 of Gill (1980) that

$$\langle M_i \rangle (t) = A_i(t)$$

and that the product $M_i(t)\, M_j(t)$ is a martingale when $i \neq j$. These results are used repeatedly in Chapter III.

A.2. Central Limit and Embedding Theorems for Martingales

Let $\{M_n(t);\ n=1, 2, \ldots\}$ be a sequence of square-integrable martingales on $[0,1]$. Let V be a continuous, increasing real function on $[0,1]$ with $V(0) = 0$. Let M be a Gaussian, independent-increments process on $[0,1]$ satisfying

$$E\, M(t) = 0 \quad \text{and} \quad E\, M^2(t) = V(t) , \quad t \in [0,1] .$$

Then the following theorem is an easy corollary of Proposition 1 of Rebolledo (1980).

Theorem A.1. Suppose that the size of the largest jump of M_n on $[0,1]$ is bounded by c_n, where $c_n \downarrow 0$ as $n\to\infty$. Suppose further that

$$\langle M_n \rangle(t) \xrightarrow{p} V(t) ,$$

as $n\to\infty$, for all $t \in [0,1]$. Then

$$M_n \xrightarrow{\mathcal{L}} M \text{ as } n\to\infty .$$

Suppose we have a discrete-time martingale difference sequence $\{X_k, F_k; k=0, 1, \ldots\}$ on $\{\Omega,F,P\}$ such that

(A.1) $$X_0 = 0$$

and

(A.2) $$|X_k| \leq B \text{ for all } k .$$

Define

(A.3) $$v_k = E(X_k^2|F_{k-1}) .$$

Let $A_1, A_2, \ldots$ be a sequence of i.i.d. random variables, uniformly distributed on $[0,1]$, on a different probability space $\{X,A,\mu\}$. Let

$$\{\Omega^*,F^*,P^*\} = \{\Omega\times X, F\times A, P\times\mu\}$$

be the product space, and let $F_k^* = F_k \times A_{2k}$, where $A_k = \sigma\{A_1, \ldots, A_k\}$. Let X_k^*, v_k^*, and A_k^* be X_k, v_k, and A_k considered as random variables on Ω^*.

Theorem A.2. (Skorokhod embedding for discrete-time martingales). There exists a standard Brownian motion $W(\cdot)$ and a sequence of random variables $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \ldots$ on $\{\Omega^*, F^*, P^*\}$ such that (A.4) holds.

(A.4a) $$X_k^* = W(\tau_k) - W(\tau_{k-1})$$

(A.4b) $$E(\tau_k - \tau_{k-1} | F_{k-1}^*) = v_k^*$$

(A.4c) $$\mathrm{var}(\tau_k - \tau_{k-1} | F_{k-1}^*) \leq 2B^2 v_k^*$$

(A.4d) τ_k is F_k^*-measurable and the pre-τ_k σ-algebra of $W(\cdot)$ is contained in F_k^*.

Remark. Theorem A.2 is very similar to the presentation on pages 90-92 in Freedman (1971). However, Freedman commits a serious error in his definition of the stopping times τ_n on page 92. See also Theorem A.1 on page 269 of Hall and Heyde (1980).

Proof. Condition on F_k^*. Arguing as in Freedman (1971), pp. 68-70, one can show that there exist nonnegative functions $U_{k+1}(\cdot,\cdot)$ and $V_{k+1}(\cdot,\cdot)$ on $\mathbb{R} \times [0,1]$ such that

(A.5) $$U_{k+1}(X_{k+1}^*, A_{2k+1}^*) = X_{k+1}^* \quad \text{for} \quad X_{k+1}^* \geq 0$$

(A.6) $$V_{k+1}(X_{k+1}^*, A_{2k+1}^*) = X_{k+1}^* \quad \text{for} \quad X_{k+1}^* \leq 0 ,$$

and

(A.7) $$\mathcal{L}(X_{k+1}^* | F_k^*, U_{k+1}, V_{k+1}) = G(U_{k+1}, V_{k+1}) .$$

Here, $G(u,v)$ is the mean 0 probability distribution with all mass on u and $-v$. Also, U_{k+1} has been written for $U_{k+1}(X^*_{k+1}, A^*_{2k+1})$, and V_{k+1} has been written for $V_{k+1}(X^*_{k+1}, A^*_{2k+1})$. The random variable A^*_{2k+1} is necessary only if the distribution of X^*_{k+1}, conditional on F^*_k, has atoms. The idea is that the distribution of X^*_{k+1}, given F^*_k, can be decomposed into a mixture of two-point distributions, each with mean 0.

Now condition on F^*_k, A^*_{2k+1}, and X^*_{k+1}. The random variable A^*_{2k+2} can be used to construct a Brownian motion which is conditioned to hit U_{k+1} before $-V_{k+1}$ if $X^*_{k+1} = U_{k+1}$ and which is conditioned to hit $-V_{k+1}$ before U_{k+1} if $X^*_{k+1} = -V_{k+1}$. (cf. Karlin and Taylor (1975), p. 378). Let

$$s_{k+1} = \inf\{s \geq 0: W_{k+1}(s) = U_{k+1} \text{ or } -V_{k+1}\} . \tag{A.8}$$

Then conditional on $\{F^*_k, U_{k+1}, V_{k+1}\}$, $W_{k+1}(\cdot)$ is a standard Brownian motion and

$$E(s_{k+1}|F^*_k, U_{k+1}, V_{k+1}) = U_{k+1} V_{k+1} . \tag{A.9}$$

Also, by Lemma (146) of Freedman (1971), p. 92,

$$E(s^2_{k+1}|F^*_k, U_{k+1}, V_{k+1}) \leq 2B^2 U_{k+1} V_{k+1} . \tag{A.10}$$

Since

$$E(U_{k+1} V_{k+1}|F^*_k) = E(X^{*2}_{k+1}|F^*_k) = v^*_k , \tag{A.11}$$

(cf. Freedman (1971), p. 70), it follows that

(A.12a) $$X^*_{k+1} = W_{k+1}\ (s_{k+1})$$

(A.12b) $$E(s_{k+1}|F^*_k) = v^*_{k+1}$$

(A.12c) $$\mathrm{var}(s_{k+1}|F^*_k) \leq 2B^2\ v^*_{k+1}\ .$$

(A.12d) The σ-algebra generated by $W_{k+1}(\cdot)$ is contained in F^*_{k+1} .

Now define the sequence $0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \ldots$ by

(A.13) $$\tau_k = \sum_{i=1}^{k} s_i\ ,$$

and define $W(\cdot)$ by

(A.14) $$W(t) = \{\sum_{i=1}^{k} W_i(s_i)\} + W_{k+1}\ (t-\tau_k)$$

for $\tau_k \leq t < \tau_{k+1}$. It is easy to see that the pre-τ_k σ-algebra of $W(\cdot)$ is contained in F^*_k, and that τ_k is F^*_k-measurable. Conditional on $\{F^*_k, U_{k+1}, V_{k+1}\}$, $W_{k+1}(\cdot)$ is a standard Brownian motion and s_{k+1} is a stopping time for this Brownian motion. Since the construction of $W(\cdot)$ just amounts to patching together stopped Brownian motions, each of which is independent of the σ-algebra generated by the previous ones, $W(\cdot)$ is itself a standard Brownian motion on $[0, \sum_1^\infty s_i)$. This, together with (A.12) proves the theorem.

Finally, let us state the continuous-time analog of Theorem A.2. Let $\{M(t), F_t;\ t \geq 0\}$ be a square-integrable martingale on $\{\Omega, F, P\}$. Suppose that the jumps of $M(\cdot)$ are of size $\leq B$.

Theorem A.3. (Skorokhod embedding for continuous-time martingales). On $\{\Omega^*,\mathcal{F}^*,P^*\}$, an enlarged version of $\{\Omega,\mathcal{F},P\}$, there exists a family of increasing random variables $\{\tau_t;\ t \geq 0\}$ and a standard Brownian motion $W(\cdot)$ such that

(A.15a) $$M^*(t) = W(\tau_t)$$

(A.15b) $$\tau_t - \langle M^* \rangle(t) \text{ is an } \mathcal{F}_t^*\text{-martingale}$$

(A.15c) $$\langle \tau_t - \langle M^* \rangle \rangle(t) \leq 2B^2 \langle M^* \rangle(t)$$

(A.15d) τ_t and the pre-τ_t σ-algebra of $W(\cdot)$ are $\mathcal{F}_t^*$-measurable.

Remark. I don't know how to prove Theorem A.3. The interested reader may wish to compare Theorem A.3 with the embedding theorems of Monroe (1972, 1978).

REFERENCES

Aalen, O. O. (1977). Weak Convergence of Stochastic Integrals Related to Counting Processes. Z. Wahrscheinlichkeitsth. verw. Geb. 38, 261-277. Correction: Vol. 48 (1979), 347.

Aalen, O. O. (1978). Nonparametric Inference for a Family of Counting Processes. Ann. Statist. 6, 701.726.

Aalen, O. O. (1980). A Model for Nonparametric Regression Analysis of Counting Processes. Springer Lecture Notes in Statistics 2, 1-25.

Andersen, P. K. and Gill, R. (1981). Cox's Regression Model for Countint Processes: A Large Sample Study. Submitted to Ann. Statist.

Armitage, P. (1975). Sequential Medical Trials, 2nd edition. Blackwell Scientific Publications, Oxford.

Bailey, K. R. (1979). The General Maximum Likelihood Approach to the Cox Regression Model. Ph.D. Dissertation, University of Chicago, Chicago, Illinois.

Cox, D. R. (1963). Large Sample Sequential Tests for Composite Hypotheses. Sankhya A 25, 5-12.

Cox, D. R. (1972). Regression Models and Life Tables (with discussion). J. Roy. Statist. Soc. B 34, 187-220.

Cox, D. R. (1975). Partial Likelihood. Biometrika 62, 269-276.

Doob, J. L. (1953). Stochastic Processes. John Wiley & Sons, New York.

Efron, B. (1977). The Efficiency of Cox's Likelihood Function for Censored Data. J. Amer. Statist. Assn. 72, 557-565.

Feller, W. (1971). An Introduction to Probability Theory and Its Applications Vol. II, 2nd edition. John Wiley & Sons, New York.

Freedman, D. (1971). Brownian Motion and Diffusion. Holden-Day, San Francisco.

Gail, M., DeMets, D., and Slud, E. (1981). Simulation Studies on Increments of the Two-Sample Log Rank Test for Survival Data with Application to Group Sequential Boundaries. Proceedings of Special Topic IMS Meeting on Survival Statistics, Columbus, Ohio, October 1981.

Gill, R. D. (1980). Censoring and Stochastic Integrals. Mathematical Centre Tracts 124, Mathematisch Centrum, Amsterdam.

Hall, P. and Heyde, C. C. (1980). Martingale Limit Theory and Its Application. Academic Press, New York.

Jones, D. and Whitehead, J. (1979). Sequential Forms of the Log Rank and Modified Wilcoxon Test for Censored Data. Biometrika 66, 105-113.

Karlin, S. and Taylor, H. M. (1975). A First Course in Stochastic Processes, 2nd edition. Academic Press, New York.

Liptser, R. S. and Shiryayev, A. N. (1977). Statistics of Random Processes I. Springer-Verlag, New York-Heidelberg-Berlin.

Liptser, R. S. and Shiryayev, A. N. (1978). Statistics of Random Processes II. Springer-Verlag, New York-Heidelberg-Berlin.

Monroe, I. (1972). On Embedding Right Continuous Martingales in Brownian Motion. Ann. Math. Statist. 43, 1293-1311.

Monroe, I. (1978). Processes that can be Embedded in Brownian Motion. Ann. Prob. 6, 42-56.

Rebolledo, R. (1980). Central Limit Theorem for Local Martingales. Z. Wahrscheinlichkeitsth. verw. Geb. 51, 269-286.

Slud, E. (1982). Sequential Linear Rank Tests for Two-Sample Censored Survival Data. Submitted to Ann. Statist.

Tsiatis, A. (1981a). A Large Sample Study of Cox's Regression Model. Ann. Statist. 9, 93-108.

Tsiatis, A. (1981b). The Asymptotic Joint Distribution of the Efficient Scores Test for the Proportional Hazards Model Calculated Over Time. Biometrika 68, 311-315.

Whitehead, J. (1978). Large Sample Sequential Methods with Application to the Analysis of 2×2 Contingency Tables. Biometrika 65, 351-356.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 20	2. GOVT ACCESSION NO.: AD-A120646	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): LARGE SAMPLE THEORY FOR SEQUENTIAL ANALYSIS OF THE PROPORTIONAL HAZARDS MODEL		5. TYPE OF REPORT & PERIOD COVERED: TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): THOMAS SELLKE		8. CONTRACT OR GRANT NUMBER(s): N00014-77-C-0306
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR-042-373
11. CONTROLLING OFFICE NAME AND ADDRESS: Office Of Naval Research Statistics & Probability Program Code 411SP Arlington, VA 22217		12. REPORT DATE: August 1982 13. NUMBER OF PAGES: 83
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report): UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

PROPORTIONAL HAZARDS MODEL, SEQUENTIAL ANALYSIS.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

PLEASE SEE REVERSE SIDE.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT # 20

SUMMARY

An appropriate large sample theory for sequential analysis of the Cox proportional hazards model is developed. For clinical trials with simultaneous entry of patients, the efficient score process of the partial likelihood is easily seen to be a martingale. It follows that, in a time scale based on the observed Fisher information, the score process and the properly normalized maximum partial likelihood estimator behave asymptotically like Brownian motion. When entry is staggered, the efficient score process is no longer a martingale in general. However, if patients in a staggered-entry clinical trial are assumed to be independent and identically distributed, independently of entry time, then the score process is well approximated by a martingale. The asymptotic results involving weak convergence to Brownian motion hold as before.

END

FILMED